Final Report

Grant NGL-05-009-079

National Aeronautics & Space Administration

APPLICATION OF SIGNAL DETECTION THEORY TO OPTICS

Carl W. Helstrom

Department of Applied Physics & Information Science

University of California, San Diego

La Jolla, Calif. 92037

Abstract

The first part of the report summarizes research carried out under this grant and now in the published literature; included are basic quantum detection and estimation theory, applications to optics, photon counting, and filtering theory. The second part describes recent work on the restoration of degraded optical images received at photoelectrically emissive surfaces; the data used by the method are the numbers of electrons ejected from various parts of the surface.

I. Survey of Published Results

The research accomplished under this grant has in large part been published in the papers listed on pages 9-11. This final report will be mostly devoted to describing our recent work on the restoration of degraded optical images received at photoelectrically emissive surfaces, but first we recapitulate our previous investigations. In particular we call attention to our review paper in Progress in Optics (§I, 3)*, which summarizes the elements of quantum detection theory and its application to the detection of light sources and the estimation of their parameters. Earlier reviews (§I, 1,2) described the state of the theory at the time, particularly with regard to its application to communications.

The function of the receiver in a communication system is to decide which of a set of known signals has been transmitted during each element of a uniform sequence of intervals; it issues its decisions as a sequence of symbols representing its best version of the transmitted message. In a radar the purpose of the receiver is to decide whether the echo from a target is present in its input and, if so, to estimate the range and velocity of the target. In optical communication and radar systems the information on which such decisions and estimates are based is embodied in the field at the aperture of the receiver, where it is corrupted by random background radiation. Because that field is governed by the laws of quantum mechanics, the methods of classical decision and estimation theories cannot be applied, and those theories must be reformulated to

---

* This notation refers to paper 3 of §I of the bibliography on pages 9-11.

take into account quantum-mechanical restrictions on measurability of the fields (§I, 3).

In quantum detection, the decisions made by the receiver are considered as choices among a set of density operators $\rho_1, \rho_2, \ldots, \rho_M$, each describing the aperture field in the presence of one of the possible transmitted signals. The density operators act in the Hilbert space $\mathscr{H}_R$ of the field. The choices are based on measurements of the field and calculations involving their outcomes, and the entire procedure can be viewed as applying to the field a resolution of the identity operator $\underset{\sim}{1}$ in $\mathscr{H}_R$ into a set of positive-definite Hermitian operators, $\Pi_1, \Pi_2, \ldots, \Pi_M$,

$$\Pi_1 + \Pi_2 + \cdots + \Pi_M = \underset{\sim}{1} ,$$

such that the probability that the receiver decides that the j-th signal was sent when really the k-th was sent is the trace $\mathrm{Tr}(\rho_k \Pi_j)$. The optimum receiver will apply that resolution of the identity for which the average probability of error is minimum. It has been shown that the operators $\Pi_1, \Pi_2, \ldots, \Pi_M$ attaining minimum error probability need not commute, but that when they do not, an ancillary apparatus can in principle be coupled with the receiver in such a way that the same minimum error probability can be attained by measuring commuting projection operators in the Hilbert space of the combination (§I, 5).

When the receiver estimates parameters of a received optical signal, such as its arrival time and Doppler shift, errors are introduced by the background radiation and by the stochastic quantum-mechanical nature of the signal itself. Lower bounds on the mean-square errors of such estimates

can be set by quantum-mechanical counterparts of the Cramér-Rao inequality of classical statistics. Two forms of the inequality are known, and sometimes the one, sometimes the other yields the superior bounds (§I, 5).

Quantum detection theory was originally formulated in terms of an ideal receiver consisting of a lossless box or cavity into which the incident light was admitted by opening an aperture during an observation interval (0,T). Measurements were then to be made on the electromagnetic field inside the cavity at a later time $t > T$ in order to decide whether it contained a component attributable to the source or object sought, or in order to estimate certain parameters, such as the radiance or direction of an optical source. Ordinarily an optical instrument bases its decisions and estimates about a source on the electromagnetic field at its aperture, for instance by focusing the incident light onto a photographic plate or a counter whose response is presented to an observer. We wanted, therefore, to eliminate that artifice of the lossless cavity and to express detection and estimation strategies directly in terms of the aperture field. This was done first for threshold detectors (§II, 1) and estimators (§II, 2). *Optimum detectors and estimators*, as well as *Cramér-Rao bounds on mean-square* errors of estimates of parameters, must in quantum mechanics be based on the density operators of the aperture field, and by decomposing that field into spatio-temporal modes it has been possible to write down the necessary density operators in a broad class of problems involving coherent or naturally incoherent light (§II, 3; §I, 3, §4 and §5).

The resolution of close point sources of incoherent light could be conveniently studied by means of this modal decomposition of the aperture

field. Two formulations were treated by quantum detection theory. In the first, an observer is to decide whether two point sources of known radiant power and position are present, or whether only a single source having twice the radiant power and located at the midpoint is present. In the second formulation either one or the other of two point sources may be radiating during the observation interval, as in a binary communication system in which 0's and 1's are sent by turning on one source or the other. In both formulations the density operators do not commute. The basic eigenvalue equation of quantum detection theory was written as an integral equation by expressing the density operators in the coherent-state representation, and by assuming the absence of background radiation it was possible to solve the integral equations for both formulations in closed form and calculate the error probabilities incurred by the optimum decision strategies (§I, 4).

The information of interest in a scene emitting or reflecting natural light is embodied in its radiance distribution as a function of position and frequency, and a camera or a telescope is essentially an instrument for estimating that radiance distribution, which the laws of light propagation translate into the spatio-temporal coherence function of the field at the aperture of the instrument. The restoration of images degraded by diffraction, aberrations, and atmospheric turbulence can be viewed as a problem of estimating a comprehensive set of parameters of the aperture field or of the radiant source producing it. Treatment of this problem by conventional linear estimation methods, assuming additive Gaussian noise, is often unjustified. Within the framework of classical optics it was

shown how the random fluctuations of the light fields cause large errors in estimates of the radiance of an object plane at points closer together than half the Rayleigh resolution distance (§II, 4). Conditions were derived under which a formula of the Shannon type holds for the information transfer from an incoherently radiating object plane (§II, 5). An analysis was made of a system that estimates the radiance distribution of an object plane by focusing it on a dense array of photoelectric detectors and processing by linear means the numbers of photoelectrons emitted by each. Equations of the Wiener type were obtained for the optimum linear processor, with the noise term embodying the classical fluctuations of the object and background light fields, and the quantum fluctuations of the photoemissive process (§II, 6).

The ideal quantum receiver of a coherent signal of random phase counts the number of photons in a single mode of the aperture field, or in a mode of the lossless cavity suitably matched to the signal. The number of photons has a Laguerre distribution in the presence of a signal and a Bose distribution in its absence. Detection and error probabilities for such an optimum receiver have been calculated (§III, 1). In practice the signal will be passed through a narrowband filter in order to cut out as much background light as possible and will then be focused on a photon counter; detection and error probabilities for such a system have also been calculated (§III, 2).

For many detectors, and particularly for the likelihood-ratio detectors prescribed by detection theory, it is not possible to calculate analytically the probability distributions of the statistics on which are based decisions

about the presence or absence of a signal, but one can work out their moment-generating functions, which are the Laplace transforms of the distributions. Methods for determining the cumulative distribution of a statistic from its moment-generating function are therefore of interest, and two have been studied, one involving an expansion in terms of Laguerre functions (§IV, 1), the other the evaluation of the inverse Laplace transform by the method of steepest descent (§IV, 3).

These techniques must be applied in analyzing the performance of a photoelectric detector proposed some time ago.* An observer is to decide whether a certain image is present or not amid background light incident on a photoelectrically emissive surface divided like a mosaic into many small, isolated elements. The decisions are based on the numbers of electrons emitted from each element, and these have Poisson distributions. The likelihood ratio is formed from the numbers and compared with a decision level chosen, for instance, to yield a pre-assigned false-alarm probability. The moment-generating functions of the logarithm of the likelihood ratio can be expressed in terms of the illuminances caused by the background and by the image sought, and by finding their inverse Laplace transforms by the method of steepest descent, false-alarm and detection probabilities have been calculated. This detector has been compared with the threshold detector† and with one that simply counts the total number of photoelectrons emitted from a circle concentric with the

---

* C. W. Helstrom, "Detection and resolution of optical signals". Trans.IEEE, vol. IT-10, pp. 275-287 (October, 1964).

† Ibid.

expected image and having optimum radius; an image with a Gaussian distribution of illuminance was postulated. It was found that the optimum likelihood-ratio processing of the photoelectrons yields only slightly greater detection probability than that of the simple counter (§III, 3).

In a thesis completed under this grant (§V, 5), Dr. C. L. Rino, now of Stanford Research Institute, developed and analyzed methods for solving the convolutional integral equation describing, inter alia, the linear filtering undergone by an image in passing through an isoplanatic optical system. It is this integral equation that must be solved in restoring degraded optical images by linear estimation. Most attention was given to one-dimensional problems. Use of the discrete Fourier transform was investigated (§V, 3,4), and the absence of data beyond the endpoints of the data interval was taken care of by using causal Wiener filters derived for the semi-infinite interval. The extrapolation of the solution beyond the ends of the data interval was studied by means of expansions in terms of prolate-spheroidal wavefunctions, and it was shown how the properties of those functions explain the large errors inherent in such extrapolations (§V, 1). These functions were also used to treat the detection and estimation of bandlimited signals (§V, 2).

Papers prepared under NASA Grant NGL-05-009-079*

I. Basic Theory

1. "Quantum detection and estimation theory", J. Statistical Physics, vol. 1, #2, pp. 231-252 (1969).
2. "Quantum-mechanical communication theory", Proc. IEEE, vol. 58, pp. 1578-1598 (October, 1970), with J. W. S. Liu and J. P. Gordon.
3. "Quantum detection theory", in Progress in Optics, E. Wolf (ed.), North-Holland Publ. Co., Amsterdam, vol. X, pp. 291-369 (1972).
4. "The resolution of point sources of light as analyzed by quantum detection theory", IEEE Transactions on Information Theory, vol. IT-19, pp. 389-398 (July, 1973).
5. "Noncommuting observables in quantum detection and estimation theory", with R. S. Kennedy, IEEE Transactions on Information Theory (in press).

II. Applications to Optics

1. "Detection of incoherent objects by a quantum-limited optical system", J. Opt. Soc. Am., vol. 59, pp. 924-936 (August, 1969, Part 1).
2. "Estimation of object parameters by a quantum-limited optical system", J. Opt. Soc. Am., vol. 60, pp. 233-242 (February, 1970).
3. "Modal decomposition of aperture fields in detection and estimation of incoherent objects", J. Opt. Soc. Am., vol. 60, pp. 521-530 (April, 1970).
4. "Resolvability of objects from the standpoint of statistical parameter estimation", J. Opt. Soc. Am., vol. 60, pp. 659-666 (May, 1970).

---

* The author of the papers in §§I-IV was C. W. Helstrom unless otherwise stated.

5. "Information transfer from incoherently radiating objects", J. Opt. Soc. Am., vol. 60, pp. 1608-1616 (December, 1970).

6. "Linear restoration of incoherently radiating objects", J. Opt. Soc. Am., vol. 62, pp. 416-423 (March, 1972).

III. Photon Counting

1. "Performance of an ideal quantum receiver of a coherent signal of random phase", IEEE Transactions on Aerospace and Electronic Systems, vol. AES-5, pp. 562-564 (May, 1969).

2. "Photoelectric detection of coherent light in filtered background light", IEEE Transactions on Aerospace and Electronic Systems, vol. AES-7, pp. 210-213 (January, 1971).

3. "Optimum detection of an optical image on a photoelectric surface", by Carl W. Helstrom and Lily Wang, IEEE Transactions on Aerospace and Electronic Systems (July, 1973).

IV. Miscellaneous

1. "Approximate calculation of cumulative probability from a moment-generating function", Proc. IEEE, vol. 57, pp. 368-369 (March, 1969).

2. "Comment on 'On the detection of a known signal in a non-Gaussian noise process'", J. Acoust. Soc. Am., vol. 45, p. 1055 (April, 1969).

3. "Numerical calculation of cumulative probability from the moment-generating function", by Lily Wang, Proc. IEEE, vol. 60, pp. 1452-1453 (November, 1972).

V. Papers by C. L. Rino on Application of Filtering Theory

1. "Bandlimited image restoration by linear mean-square estimation", J. Opt. Soc. Am., vol. 59, pp. 547-553 (May, 1969).

2. "The application of prolate spheroidal wave functions to the detection and estimation of bandlimited signals", Proc. IEEE, vol. 58, pp. 248-249 (February, 1970).

3. "The inversion of covariance matrices by finite Fourier transforms", IEEE Transactions on Information Theory, vol. IT-16, pp. 230-232 (March, 1970).

4. "Factorization of spectra by discrete Fourier transforms", IEEE Transactions on Information Theory, vol. IT-16, pp. 484-485 (July, 1970).

5. "Discrete transform methods in solving the instrument convolutional integral equation in filtering stationary time series", Thesis, Department of Applied Physics & Information Science, University of California, San Diego (1970).

VI. Other Theses

1. L. Wang, "Detection and Estimation of an Optical Image by Photon-Counting Techniques", Department of Applied Physics & Information Science, University of California, San Diego (June, 1973).

2. Y.-M. Hong, "Optical Signal Processing: Poisson Image Restoration and Shearing Interferometry", Department of Applied Physics & Information Science, University of California, San Diego (September, 1973).

II. RESTORATION OF DEGRADED OPTICAL IMAGES*

1. Maximum-Likelihood Image Restoration

The restoration of degraded optical images has in the past been treated largely as linear filtering of a spatial random process in the presence of additive Gaussian noise.[1-5] It is only under special conditions that this is an accurate model of the image. Although the electromagnetic fields of the light from the object and the background combine additively, it is not the field of the object light that is of interest, but the radiance distribution of the object or scene, which is related to the covariance function of the field. Furthermore, what is measured at the image plane of an optical system is the illuminance of the incident light, which is also a quadratic functional of the field. When photographic film is used, the noise depends on the illuminances due to object and background in a complicated way. The simple additivity of the Gaussian model is seldom realistic.

In an endeavor to study image restoration in terms of a more realistic, yet mathematically tractable model, we have postulated that the incident light is focused on a photoelectrically emissive surface, which is divided like a mosaic into a number N of small, insulated regions A_i of area ΔA. During an observation interval of duration T the numbers n_i of photoelectrons ejected from each of these elements A_i are recorded and constitute the data on the basis of which an estimate of the undegraded image is formed.[6] The system is illustrated in Fig. 1.1.

Let the illuminance that would appear at the point $\underset{\sim}{x}$ of the image plane, were the undistorted geometrical image present, be $B(\underset{\sim}{x})$; this is proportional to

---

*The research described in this part was carried out in collaboration with Y.-M. Hong and is described in his doctoral thesis, "Optical Signal Processing: Poisson Image Restoration and Shearing Interferometry", Dep't. of Applied Physics & Information Science, University of California, San Diego; September, 1973.

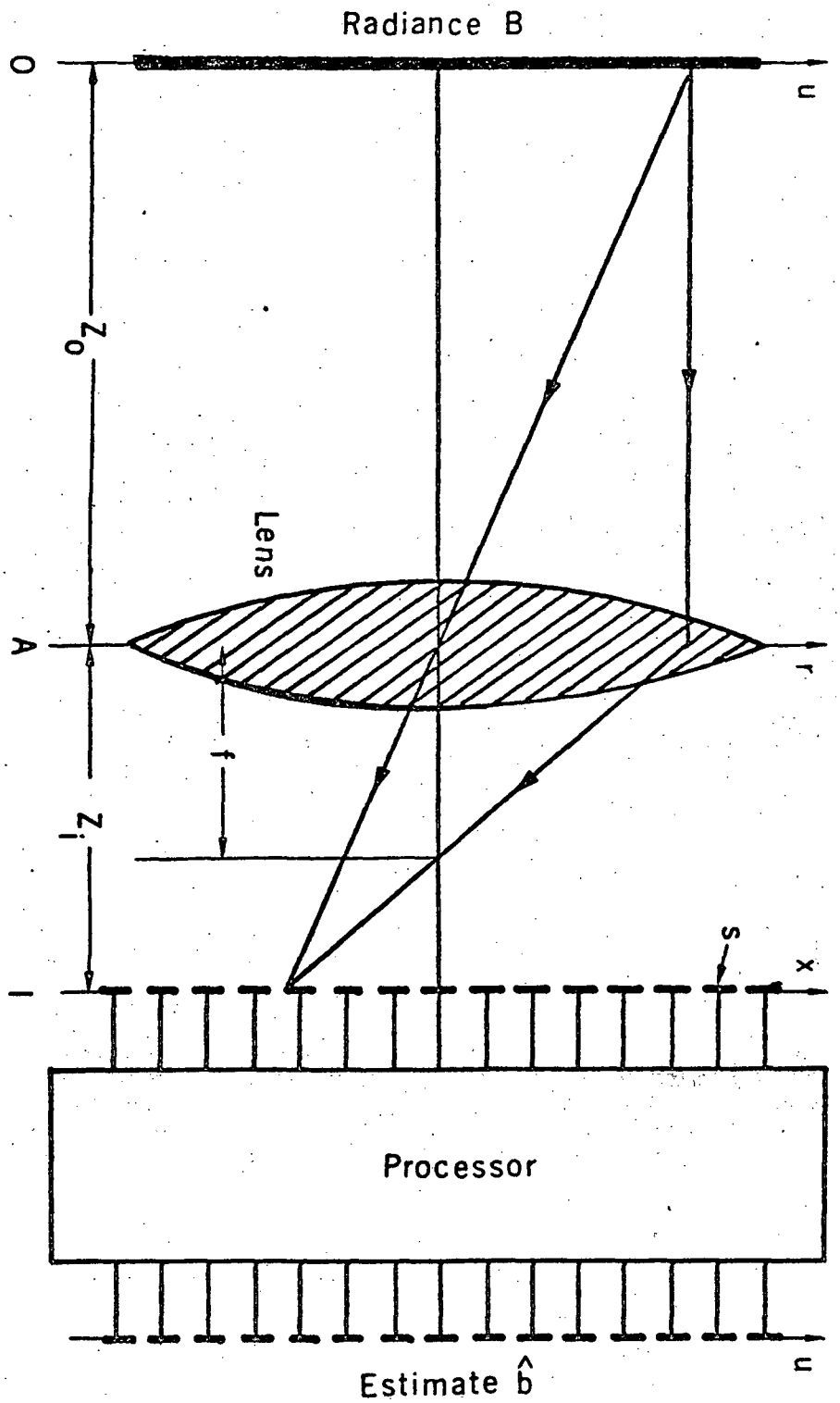

Fig. 1.1 Image restoration system. The elements s of area ΔA emit numbers n_i of electrons that are processed to yield estimates $\hat{B} = B_0 + \hat{b}$ of samples of the true image of the radiance distribution B. O = object plane, A = aperture, I = image plane. A narrowband spectral filter for object and background light is not shown.

the radiance distribution of the object plane with its dimensions scaled according to the magnification factor of the optical system. The actual illuminance at point $\underset{\sim}{x}$ is

$$I(\underset{\sim}{x}) = \int K(\underset{\sim}{x},\underset{\sim}{u})\, B(\underset{\sim}{u}) d^2\underset{\sim}{u}, \tag{1.1}$$

where $K(\underset{\sim}{x},\underset{\sim}{u})$ is the incoherent point spread function of the optical system and the transmission medium combined. It accounts for turbulent distortions of the refractive index, diffraction at the aperture of the observing instrument, aberrations in lenses, and relative motion of object and instrument. We suppose the kernel $K(\underset{\sim}{x},\underset{\sim}{u})$ to be known.

Let $\underset{\sim}{x}_i$ be the coordinates of the center of the i-th element of the image plane, which element we suppose small enough that $I(\underset{\sim}{x})$ is effectively constant over it. Then the mean number of photoelectrons emitted by the i-th element A_i is

$$\bar{n}_i = \alpha\, I(\underset{\sim}{x}_i), \qquad i = 1,2,..,n, \tag{1.2}$$

with

$$\alpha = T\Delta A \int \eta(\nu)\Phi(\nu)(h\nu)^{-1} d\nu, \tag{1.3}$$

where $\eta(\nu)$ is the quantum efficiency of the surface at frequency ν, $\Phi(\nu)$ is the relative spectral density of the light,

$$\int \Phi(\nu) d\nu = 1,$$

and $h\nu$ is the energy of a quantum of light at frequency ν. When as we assume here the duration T is much greater than the reciprocal W^{-1} of the bandwidth of the light, $WT >> 1$, the numbers n_i of emitted photoelectrons have Poisson distributions and are statistically independent,[7]

$$\Pr\left\{n_1,\ldots,n_N\right\} = \prod_{i=1}^{N} \frac{\bar{n}_i^{\,n_i}}{n_i!} \exp\left(-\bar{n}_i\right) \tag{1.4}$$

We suppose that

$$B(\underset{\sim}{u}) = B_0 + b(\underset{\sim}{u}), \tag{1.5}$$

where B_0 is known and $b(\underset{\sim}{u})$ is to be estimated at a number M of uniformly spaced points $\underset{\sim}{u}_i$, in the intervals between which $B(\underset{\sim}{u})$ changes in a smooth manner. With

$$I_0 = \int K(\underset{\sim}{x},\underset{\sim}{u})\, B_0 d^2\underset{\sim}{u} \tag{1.6}$$

assumed independent of $\underset{\sim}{x}$, we put for the mean values of the data n_i,

$$\bar{n}_i = \alpha(I_0 + \eta_i), \tag{1.7}$$

and

$$\eta_i = \sum_{j=1}^{M} K_{ij} b_j, \qquad i = 1,2,\ldots,N, \tag{1.8}$$

where

$$K_j = K(\underset{\sim}{x}_i,\underset{\sim}{u}_j)\, \Delta A. \tag{1.9}$$

Thus in this approximation

$$I_0 = \sum_{j=1}^{M} K_{ij}\, B_0. \tag{1.10}$$

The data are the N numbers n_i distributed as in (1.4); the estimanda are the sample values b_j of the deviation of the geometrical image from a mean level B_0.

In order to apply the principle of maximum likelihood, some statistical description must be formulated for the class of expected true images $B(\underset{\sim}{u})$. We consider them as spatial Gaussian random processes with mean zero and auto-

covariance function[6]

$$\varphi(\underset{\sim}{u}_1-\underset{\sim}{u}_2) = \underset{\sim}{E}\left[B(\underset{\sim}{u}_1)B(\underset{\sim}{u}_2)\right] - B_0^{\,2} = \underset{\sim}{E}\left[\, b(\underset{\sim}{u}_1)b(\underset{\sim}{u}_2)\right]. \tag{1.11}$$

The images are spatially stationary, and the width of $\varphi(\underset{\sim}{u})$ represents the size of typical details in the object plane; the ratio

$$C = \frac{\varphi(0)}{B_0^{\,2} + \varphi(0)} \tag{1.12}$$

specifies a mean-square contrast. The Fourier transform of the autocovariance function $\varphi(\underset{\sim}{u})$ describes the distribution of spatial frequencies in objects of the class. Although the actual objects will not in general resemble what we think a Gaussian stochastic process looks like, a method that works well for an ensemble of such processes should usually be effective in restoring an object of similar structure. The structure of a scene can seldom be specified in terms less rudimentary than the average contrast and the size of typical details. The form chosen for the autocovariance function in (1.11) represents the demands we are placing on our restoration scheme. If $\varphi(\underset{\sim}{u})$ is taken to be very narrow, we are forcing the method to try to reproduce fine details, and we can expect the resulting image to be noisy. If $\varphi(\underset{\sim}{u})$ is too broad, the resulting image will be in error because details have been smeared out.

The principle of maximum likelihood states that the best estimate of the image is the one for which the joint conditional probability density function (p.d.f.) of the estimanda b_j, given the data n_i, is maximum. It is given by

$$\Pr\left(\{b_j\}\,\middle|\,\{n_i\}\right) = q\left(\{b_j\}\right)\Pr\left(\{n_i\}\,\middle|\,\{b_j\}\right) \Big/ \Pr\left(\{n_i\}\right), \tag{1.13}$$

where

$$q\left(\left\{b_j\right\}\right) = C_q \exp\left(-\frac{1}{2}\sum_{j=1}^{M}\sum_{k=1}^{M}\mu_{jk}\, b_j\, b_k\right) \tag{1.14}$$

is the prior p.d.f. of the estimanda, $\Pr\left(\left\{n_i\right\}\middle|\left\{b_j\right\}\right)$ is the product of Poisson distributionsgiven in (1.4), and

$$\Pr\left(\left\{n_i\right\}\right) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} q\left(\left\{b_j\right\}\right)\Pr\left(\left\{n_i\right\}\middle|\left\{b_j\right\}\right)\, db_1..db_M$$

is the unconditional distribution of the data. In (1.14) the μ_{jk} are the elements of a matrix $\underset{\sim}{\mu} = \underset{\sim}{\varphi}^{-1}$ that is the inverse of the covariance matrix $\underset{\sim}{\varphi}$ whose elements are

$$\varphi_{ij} = \varphi(\underset{\sim}{u}_i - \underset{\sim}{u}_j); \tag{1.15}$$

C_q in (1.14) is a normalizing constant. Maximizing the conditional p.d.f. in (1.13) is equivalent to minimizing its negative logarithm

$$F\left(\left\{b_j\right\}\right) = \sum_{i=1}^{N}(\bar{n}_i - n_i\, \ell n\, \bar{n}_i) + \frac{1}{2}\sum_{j=1}^{M}\sum_{k=1}^{M}\mu_{jk}b_j b_k, \tag{1.16}$$

from which terms independent of the estimanda $\left\{b_j\right\}$ have been dropped. The constraints on the minimization are embodied in eqs. (1.7) and (1.8), which take the distortion of the image into account.

If the numbers n_i are large and the contrast is small, the distributions of the n_i's are approximately Gaussian with mean values $\bar{n}_i$ given by (1.7) and variances $\bar{n}_i \cong \alpha I_0$. The quantity to be minimized is then, instead of (1.16),

$$F'\left(\left\{b_j\right\}\right) = \frac{1}{2}(\alpha I_0)^{-1}\sum_{i=1}^{N}(n_i - \bar{n}_i)^2 + \frac{1}{2}\sum_{j=1}^{M}\sum_{k=1}^{M}\mu_{jk}b_j b_k, \tag{1.17}$$

which must also be taken with (1.7) and (1.8). The minimizing values of the estimanda b_j are the solution of the linear equations

$$\sum_{k=1}^{M} \left(I_0 \mu_{jk} + \alpha \sum_{i=1}^{N} K_{ik} K_{ij} \right) b_k = \sum_{i=1}^{N} (n_i - \alpha I_0) K_{ij}; \tag{1.18}$$

they provide the usual least-squares estimate, which can be written in matrix form as

$$\hat{\underset{\sim}{b}} = \left(I_0 \underset{\sim}{\mu} + \alpha \underset{\sim}{K}^T \underset{\sim}{K} \right)^{-1} \underset{\sim}{K}^T (\underset{\sim}{n} - \alpha I_0 \underset{\sim}{1})$$

$$= \left(I_0 \underset{\sim}{I}_M + \alpha \varphi \underset{\sim}{K}^T \underset{\sim}{K} \right)^{-1} \varphi \underset{\sim}{K}^T \left(\underset{\sim}{n} - \alpha I_0 \underset{\sim}{1} \right), \tag{1.19}$$

where $\underset{\sim}{n}$ is an N-dimensional column vector of the data, $\underset{\sim}{1}$ a column vector of N 1's, and $\underset{\sim}{I}_M$ is the MxM identity matrix. The matrix $\underset{\sim}{K} = ||K_{ij}||$ is NxM and $\underset{\sim}{K}^T$ is its transpose. This least-squares estimate forms the starting point of our search for the maximum-likelihood estimate, which minimizes $F\left(\left\{b_j\right\}\right)$ in (1.16).

Having neither prior knowledge nor preconceptions about the class of images to be restored is equivalent statistically to attributing an infinite variance to the estimanda $\left\{b_j\right\}$. The second term of (1.16) then disappears, and one minimizes only

$$F''\left(\left\{b_j\right\}\right) = \sum_{i=1}^{N} \left(\bar{n}_i - n_i \ln \bar{n}_i \right) \tag{1.20}$$

in combination with the constraints in (1.7) and (1.8). For the least-squares estimate this leads to the simple "inverse filtering" specified by the equations

$$\hat{\underset{\sim}{b}} = \left(\alpha \underset{\sim}{K}^T \underset{\sim}{K} \right)^{-1} \underset{\sim}{K}^T \left(\underset{\sim}{n} - \alpha I_0 \underset{\sim}{1} \right). \tag{1.21}$$

We shall find that the image restorations based on prior assumptions about the class of true images, through (1.16), are seldom much more accurate than those obtained, by maximizing (1.20), without utilizing such assumptions.

2. The Method of Steepest Descents

The most successful method we tried for computing the maximum-likelihood estimate $\hat{b}_j$ of the true image was the method of steepest descents applied to the minimization of the function $F\left(\left\{b_j\right\}\right)$ in eq. (1.16).[8] Let a given set of trial values of the estimates $\hat{b}_j$ be denoted by the superscript ^{0}. We expand the function $F\left(\left\{b_j\right\}\right)$ in an M-dimensional Taylor series about the point $\left\{b_j{}^0\right\}$,

$$F\left(\left\{b_j\right\}\right) = F\left(\left\{b_j{}^0\right\}\right) + \sum_{k=1}^{M} \left(\frac{\partial F}{\partial b_k}\right)^0 \left(b_k - b_k{}^0\right)$$

$$+ \frac{1}{2} \sum_{k=1}^{M}\sum_{m=1}^{M} \left(\frac{\partial^2 F}{\partial b_k \partial b_m}\right)^0 \left(b_k - b_k{}^0\right)\left(b_m - b_m{}^0\right) \qquad (2.1)$$

and treat this as a quadratic form to be minimized. The next trial point lies in the direction of the gradient from $\underset{\sim}{b}^0$,

$$b_j = b_j{}^0 - \lambda\, g_j, \qquad (2.2)$$

where

$$g_j = \left(\frac{\partial F}{\partial b_j}\right)^0 = \alpha \sum_{i=1}^{N} K_{ij}\left(1 - \frac{n_i}{\bar{n}_i}\right) + \sum_{k=1}^{M} \mu_{jk}\, b_k{}^0, \quad j = 1,2,\ldots,M, \qquad (2.3)$$

are the components of the gradient vector at point $\underset{\sim}{b}^0$, in which we must put, from (1.7) and (1.8),

$$\bar{n}_i = \alpha\left(I_0 + \eta_i{}^0\right), \qquad (2.4)$$

$$\eta_i{}^0 = \sum_{j=1}^{M} K_{ij} b_j{}^0. \qquad (2.5)$$

The parameter λ determines how far along the gradient one proceeds, and it is determined by substituting (2.2) into (2.1) and minimizing with respect to λ,

$$\lambda = \sum_{k=1}^{M} g_k^2 \Big/ \sum_{k=1}^{M}\sum_{m=1}^{M} a_{km} g_k g_m , \tag{2.6}$$

$$a_{km} = \left(\frac{\partial^2 F}{\partial b_k \partial b_m} \right)^0 = \mu_{km} + \sum_{i=1}^{M} \frac{n_i K_{ik} K_{im}}{\left(I_0 + \eta_i^0\right)^2} . \tag{2.7}$$

The quadratic form in the denominator of λ can be written

$$\sum_{k=1}^{M}\sum_{m=1}^{M} a_{km} g_k g_m = \sum_{k=1}^{M}\sum_{m=1}^{M} \mu_{km} g_k g_m + \sum_{i=1}^{M} \frac{n_i r_i^2}{\left(I_0 + \eta_i^0\right)^2} , \tag{2.8}$$

where

$$r_i = \sum_{k=1}^{M} K_{ik}\, g_k . \tag{2.9}$$

When the trial point changes from $\underset{\sim}{b}^0$ to $\underset{\sim}{b}$ by (2.2), η_i in (2.5) changes from η_i^0 to $\eta_i^0 - \lambda r_i$. The only MxM matrix that needs to be stored in the computer is $\underset{\sim}{\mu} = \underset{\sim}{\varphi}^{-1}$, the inverse covariance matrix, and for the covariance matrix $\underset{\sim}{\varphi}$ we used, $\underset{\sim}{\mu}$ was very simple. The rest of the operations in calculating λ and the new values of $\{b_j\}$ and $\{\eta_j\}$ can be carried out with one-dimensional arrays.

Because the true image must be positive, values of b_j that sank below $-B_0$ at any stage of the iteration were set equal to $-B_0+\varepsilon$, where ε was a small positive constant. The iterations were stopped when the ratio

$$\sum_{j=1}^{M} |b_j - b_j^0| \Big/ \sum_{j=1}^{M} |b_j|$$

fell below 10^{-3}, where $\{b_j^0\}$ were the values at the beginning, $\{b_j\}$ those at the end of an iteration.

3. Generation of Trial Objects and Images

The method of steepest descents was tried out on the restoration of one-dimensional objects distorted by motion blur and diffraction. Two types of objects were used, a deterministic one composed of Gaussian peaks and random ones generated by an autoregressive process. The former had the form

$$b_i = 90\left\{ \exp\left[-(i-5)^2/4\right] + \exp\left[-(i-9)^2/4\right] + \exp\left[-(i-16)^2/4\right]\right\}. \quad (3.1)$$

Its peaks are two units in width and located at coordinates 5, 9, and 16 in a total range of twenty points.

The random object was generated by the formula

$$b_{i+1} = r\, b_i + z_{i+1}, \quad i = 1, 2, \ldots, M-1, \quad (3.2)$$

starting from an initial value b_1 that was a Gaussian random variate with mean zero and variance σ_b^2. The z_i's were independent Gaussian random variates with mean zero and variance $\sigma_b^2(1-r^2)$, as a result of which all the image variates were Gaussian with mean zero and variance σ_b^2, but with a correlation matrix given by

$$\varphi_{ij} = \sigma_b^2\, r^{|i-j|}. \quad (3.3)$$

The inverse $\underset{\sim}{\mu}$ of this matrix is especially simple. Its diagonal and first off-diagonal elements are

$$\begin{aligned} \mu_{11} &= \mu_{MM} = \sigma_b^{-2}/(1-r^2), \\ \mu_{ii} &= \sigma_b^{-2}(1+r^2)/(1-r^2), \quad i \neq 1 \text{ or } M, \\ \mu_{i,i+1} &= \mu_{i+1,i} = -\sigma_b^{-2}\, r/(1-r^2). \end{aligned} \quad (3.4)$$

All other elements of the matrix $\underset{\sim}{\mu}$ vanish.

The Gaussian variates z_i were generated from pairs of uniformly distributed random variates R_1 and R_2 in the interval (0,1) by means of the transformation

$$z = \sigma_z(-2\ \ell n\ R_1)^{1/2} \cos\ (2\pi R_2),\ \sigma_z = (\text{Var } z)^{1/2} \tag{3.5}$$

Values of b_i generated by (3.2) that fell below $-B_0$ were set equal to $-B_0$ in order to produce an object with nonnegative radiance $B_0 + b_i$. A typical object generated by this process is illustrated by the heavy line in Fig. 3.1. Its parameters are $B_0 = 3$, $r = 0.6$, $\sigma_b = 25$.

When the image is degraded by relative motion between object and recording instrument, the point spread function $K(\underset{\sim}{x},\underset{\sim}{u})$ will be constant for a range of values of $\underset{\sim}{x}$ about $\underset{\sim}{u}$ and zero elsewhere. Sampling then produces an NxM matrix $\underset{\sim}{K}$ of the form

$$K_{ij} = D^{-1},\ i = j,j+1,\ldots,j+D-1,\ j = 1,2,\ldots,\ M, \tag{3.6}$$

where $N = M+D-1$, and D is an integer proportional to the relative velocity of object and instrument; we took D=4. Although (3.6) appears to violate (1.10), we can assume that the actual object extends to the left and right of the range of sampled values and has in those margins a constant radiance B_0; the remaining analysis, with $I_0 = B_0$, is unaltered.

For diffraction the point-spread matrix was taken as

$$K_{ij} = C\left[\sin^2 W(i-j)\right] / (i-j)^2 = e_{i-j},\ |i-j|<J,$$
$$K_{ij} = 0\ ,\ |i-j| > J,$$
$$C=\left(\sum_{i=-J}^{J} e_i\right)^{-1}. \tag{3.7}$$

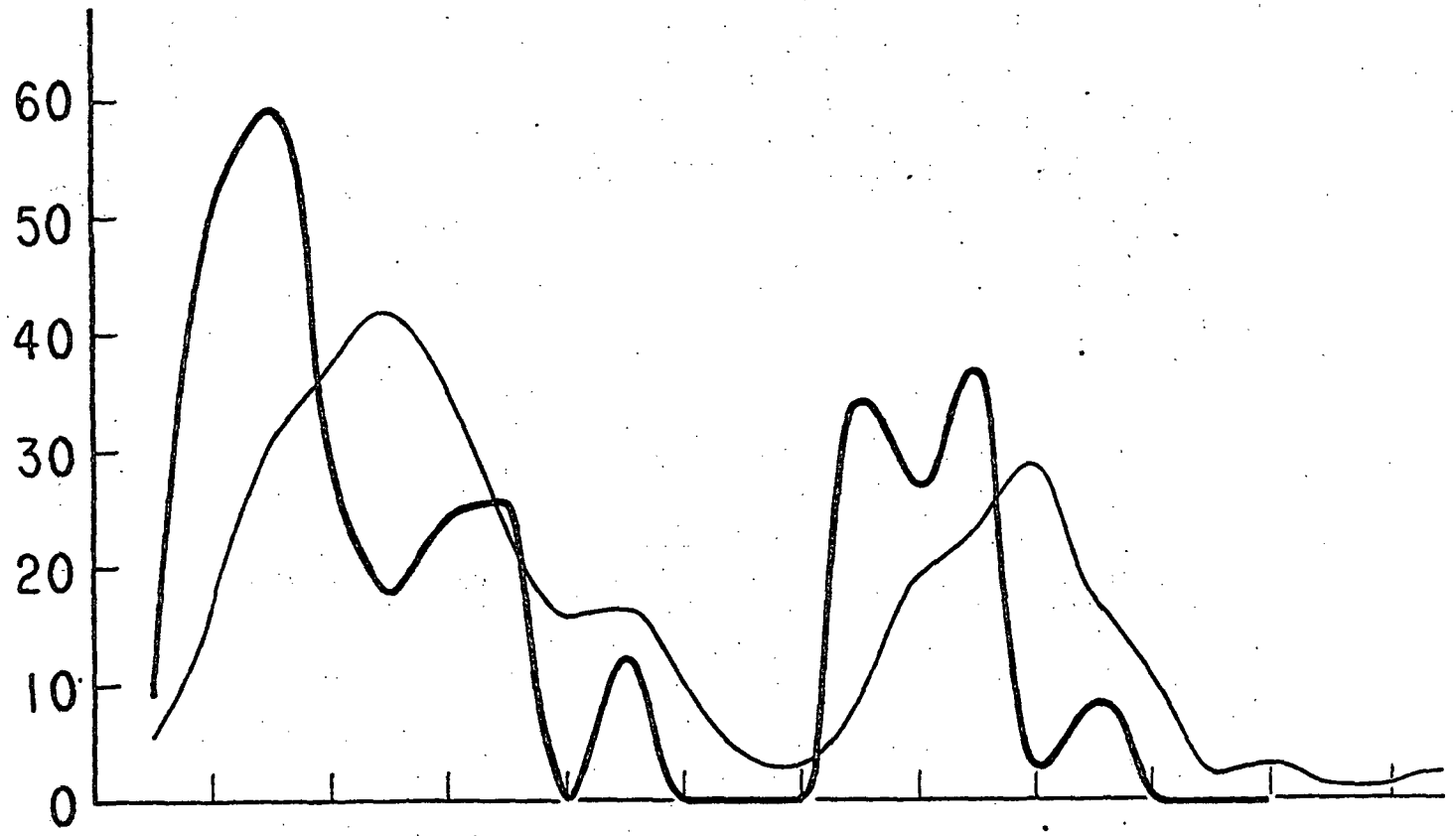

Fig. 3.1 True image (thick line) and the noisy degraded image (thin line) after blurring by relative motion. $I_0 = 3$, $a = 10$, $\sigma = 25$, $r = 0.6$.

Again we suppose that the object has wide enough margins with $b(\underset{\sim}{u}) \equiv 0$, in which its radiance is constant and equal to $B_0 = I_0$. Here W is related to the width of the aperture, which is taken to be a slit perpendicular to the one dimension of the object.

The data n_i, representing the numbers of photoelectrons ejected by the incident light from the several elements of the image surface, were generated as Poisson-distributed random variates with mean values

$$\bar{n}_i = \alpha\,(I_0 + \eta_i),$$

$$\eta_i = \sum_{j=1}^{M} K_{ij} b_j, \quad i = 1,\ldots,N. \tag{3.8}$$

Let $\lambda = \bar{n}_i$ be a typical such mean value. A pseudorandom-number program produces a uniformly distributed random variate y in (0,1), which is multiplied by e^{λ}. The sums

$$s_n = \sum_{j=0}^{n} \lambda^j/j!$$

are accumulated, starting with $s_0 = 1$, until s_n just exceeds ye^{λ}, whereupon the accumulation stops and the datum n_i is set equal to the final integer n. The probability that n_i takes on a value n is then, with $s_{-1} = 0$,

$$\Pr\left\{n_i = n\right\} = \Pr\left\{s_{n-1} < ye^{\lambda} < s_n\right\} = \lambda^n\, e^{-\lambda}/n!,$$

as required.

For mean values $\bar{n}_i$ exceeding 25 we approximated the Poisson distribution as a Gaussian with mean value $\bar{n}_i$ and variance $\bar{n}_i$; a Gaussian random variate was generated as in (3.5) with $\sigma_z = \bar{n}_i^{1/2}$, $\bar{n}_i$ was added, and the result was rounded

to the nearest integer.

The object shown as the heavy line in Fig. 3.1 was blurred by applying the matrix $\underset{\sim}{K}$ in (3.6) with D=4, and Poisson variates n_i were generated from the resulting mean values $\bar{n}_i$ obtained as in (1.8), with $\alpha = 10$. The n_i's were then divided by α to reduce them to the same scale as the object and plotted on Fig. 3.1 as the thin wavy line, which represents the recorded data on which the restoration was based.

4. The Computer Simulations

The image data n_i generated as we have just described were subjected to three image-restoration procedures and the results were compared. As a simple measure of the quality of a restoration we chose the relative average squared error, in percent, defined by

$$\mathscr{E} = 100 \times \sum_{j=1}^{M} \left(\hat{b}_j - b_j\right)^2 \Big/ \sum_{j=1}^{M} b_j^{\,2} , \tag{4.1}$$

where b_j are the known illuminance values of the true image and $\hat{b}_j$ are the estimates.

The first restoration procedure was simple inverse linear filtering, which corresponds to the least-squares estimate in the limit $\varphi(0) \to \infty$; no prior knowledge of the class of objects is presumed. It is determined from eq. (1.21). The second procedure was the method of linear least-squares estimation, as given by eq. (1.19). The resulting estimate $\hat{\underset{\sim}{b}}$ served as the starting point for seeking the nonlinear maximum-likelihood estimate, and for this the method of steepest descents described in Section 2 was applied. For all three procedures any estimates $\hat{b}_j$ lying below $-B_0$ were set equal to $-B_0$ before the average squared error $\mathscr{E}$ was calculated.

Figure 4.1 shows the results of applying these three procedures to restoring the image depicted in Fig. 3.1. As the signal-to-noise ratio $\alpha\sigma_b^{\,2}/B_0$ is large in this case, the linear least-squares estimate is not much different from the one obtained by inverse filtering as in (1.21); the relative percentage errors are 10.1% and 12.5%, respectively. The maximum-likelihood estimate, on the other hand, yields $\mathscr{E} = 2.53\%$ and lies much closer to the true image than either of the others.

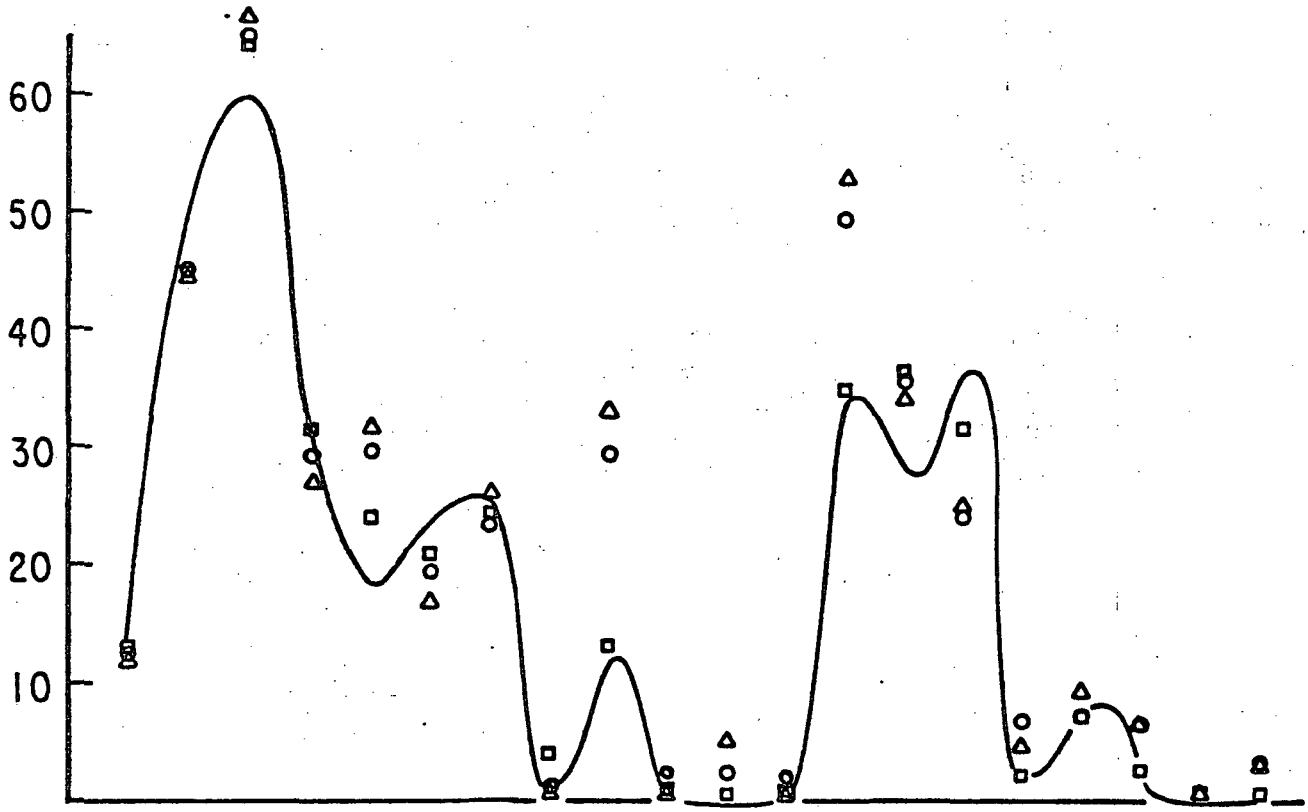

Fig. 4.1 True image of Fig. 3.1 and its restorations: o = linear estimate, Δ = inverse-filtering estimate, □ = maximum-likelihood estimate.

When establishing an image-restoration system it is often difficult to judge what covariance function $\varphi(\underset{\sim}{u})$ most aptly describes the class of true images to be restored. The robustness of the system to the use of inappropriate covariances is therefore of interest. In order to investigate it, we have tried our methods with values of the object variance σ_b^2 and the correlation coefficient r in (3.3) and (3.4) different from those with which the object in Fig. 3.1 was generated. The resulting error percentages $\mathscr{E}$ are plotted in Fig. 4.2. Inverse filtering is independent of the object covariance. The maximum-likelihood method is relatively insensitive to the value of the correlation coefficient r until it exceeds about 0.9, after which the method breaks down and yields larger errors than the linear least-squares estimator. It may be that a value of r too close to 1 imposes too much correlation on the estimated radiance values and prevents the search procedure from accommodating to the data.

The absence of any prior assumptions about the class of images being restored is equivalent to the assignment of infinite variance to the prior p.d.f. $q\left(\left\{b_j\right\}\right)$ in (1.14) and eliminates the second term from the function $F\left(\left\{b_j\right\}\right)$ to be minimized, eq. (1.16). The steepest-descent procedure minimizing (1.20) is therefore simpler and cheaper. Fig. 4.3 shows the resulting restoration of the image in Fig. 3.1. Its relative squared error amounts to 2.69%, only slightly larger than for the restoration that used the known values of σ_b^2 and r and is depicted in Fig. 4.1.

The variability of our restorations was studied by starting with the same distorted image, that is, with the same mean values $\bar{n}_i$, and generating independent sets of Poisson-distributed data n_i by starting the random number procedure with different "initializing constants." The first example used an autoregressive object with $\sigma_b = 15$, $r = 0.2$, $I_0 = 3$, and $\alpha = 10$, distorted by motion blur introduced by the matrix $\underset{\sim}{K}$ of (3.6) with D=4. Table I lists the

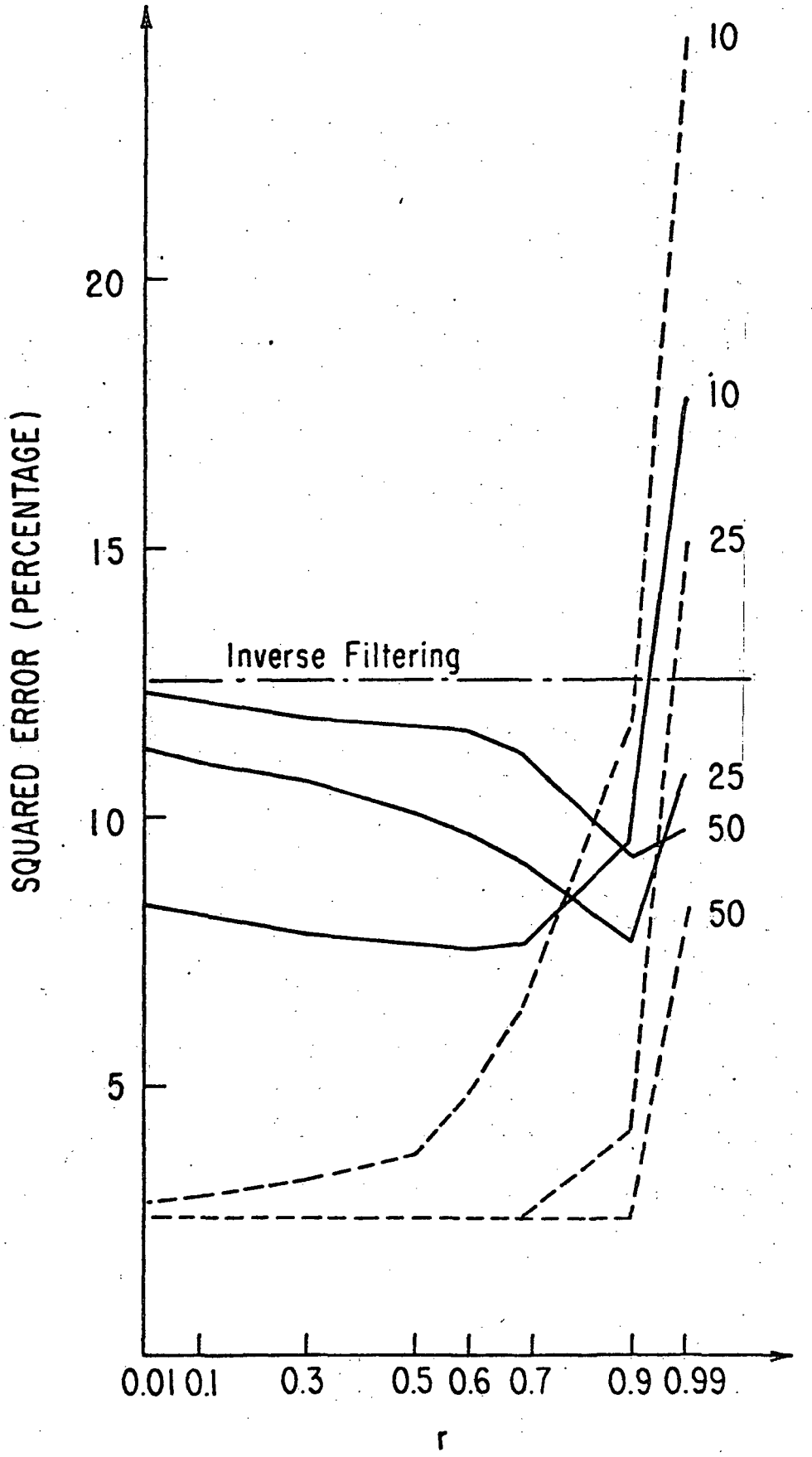

Fig. 4.2 Percentage squared error in restoring Fig. 3.1 as function of assumed correlation coefficient r; curves are indexed by assumed standard deviation σ_b. Solid lines-linear estimate; dashed lines—maximum-likelihood estimate.

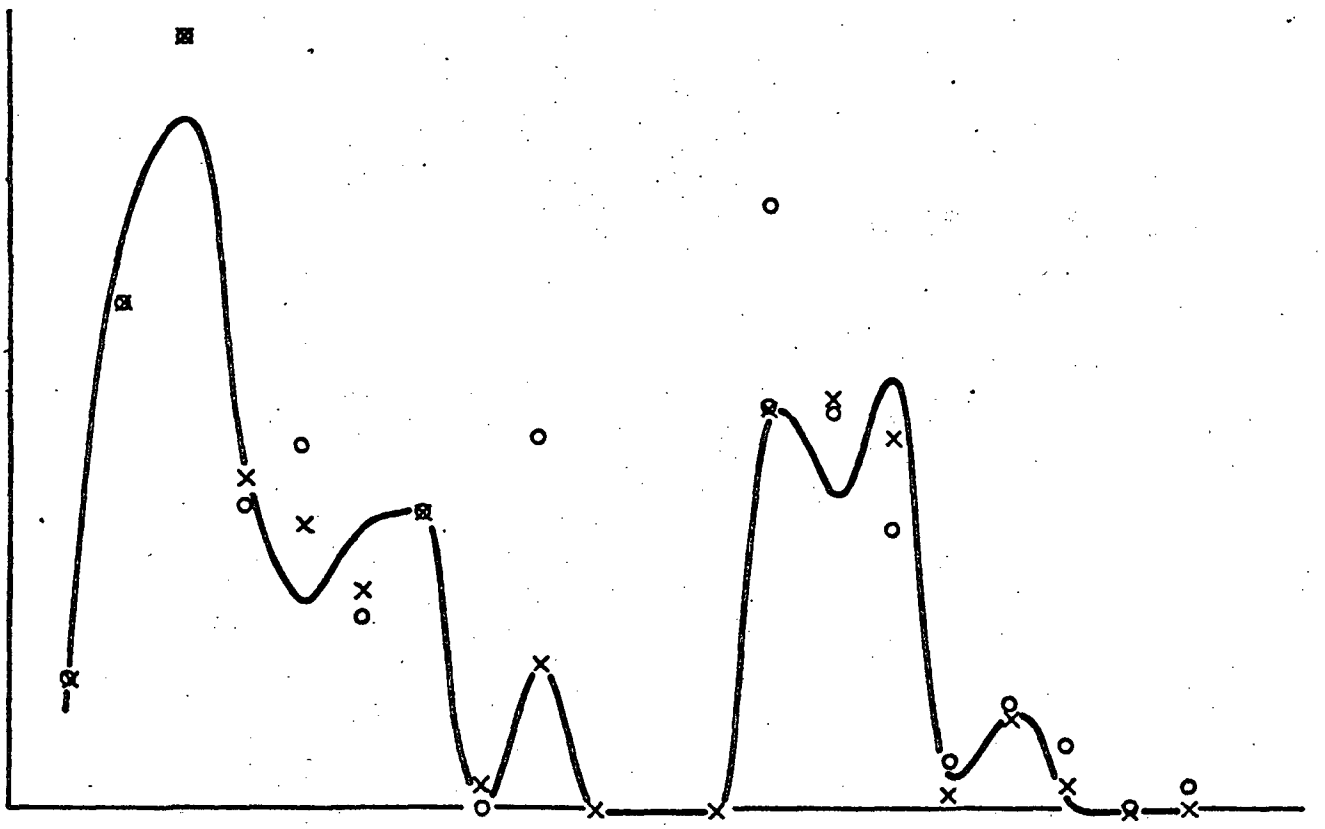

Fig. 4.3 Restoration of Fig. 3.1 without prior assumptions ($\sigma_b = \infty$). o = linear inverse filtering, × = maximum-likelihood.

Table I

$\sigma_b = 15, \quad r = 0.2, \quad I_0 = 3, \quad \alpha = 10$

Relative Squared Errors (percent)

	$\sigma_b = 15$, $r = 0.2$		$\sigma_b \to \infty$	
	linear	nonlinear	linear	nonlinear
1.	5.53	1.80	8.81	0.862
2.	2.13	2.46	1.02	1.08
3.	4.90	2.93	4.76	2.21
4.	5.95	4.28	8.20	2.44
5.	7.61	5.20	5.17	3.43
*6.	5.83	5.79	3.43	3.58
7.	8.49	4.23	8.25	3.68
8.	6.02	5.64	3.97	3.69
9.	5.49	3.67	9.77	4.76
10.	9.84	6.22	8.69	5.17
11.	7.84	6.36	7.24	5.63
12.	12.59	9.39	12.36	9.56
mean	6.85	4.83	6.81	3.84
standard deviation	2.67	2.07	3.18	2.33

*Illustrated in Figs. 4.4 and 4.5

resulting squared errors; the left-hand columns are for a restoration procedure using the correct prior variance $\sigma_b = 15$ and the correct correlation coefficient r=0.2. The right-hand columns are for $\sigma_b \to \infty$, i.e., no prior knowledge. The linear and the maximum-likelihood estimators were used; for $\sigma_b \to \infty$ the linear estimator corresponds to inverse filtering as in (1.21). The starred line refers to the restorations depicted in Figs. 4.4 and 4.5.

In ten out of twelve of these trials the smaller relative error $\mathscr{E}$ was obtained for $\sigma_b \to \infty$, i.e., by discarding all prior knowledge of the class of objects. In two cases the linear estimator was more effective than the maximum-likelihood estimator. In general, however, the maximum-likelihood estimator was rather better than the linear estimator, and as the figures for standard deviation at the bottom of the table demonstrate, the former is the more consistent.

A second set of trials with a different object of the same type yielded the squared errors listed in Table II; the starred case is illustrated in Figs. 4.6 and 4.7. Here there was a greater variability, and in all but one case the maximum-likelihood estimate was closer than the linear one to the true image. Omitting the prior information led to closer restorations in seven out of the twelve trials. Again the maximum-likelihood estimator was the more consistent, and on the average considerably better than the linear estimator.

The Gaussian "deterministic" object of eq. (3.1) and its motion-blurred image are shown in Fig. 4.8; again the matrix $\underset{\sim}{K}$ in (3.6) was used with D=4. The dip between the two left-hand peaks has been wiped out. Data were generated from this object by taking $B_0 = I_0 = 5$ and $\alpha = 3$. It was then

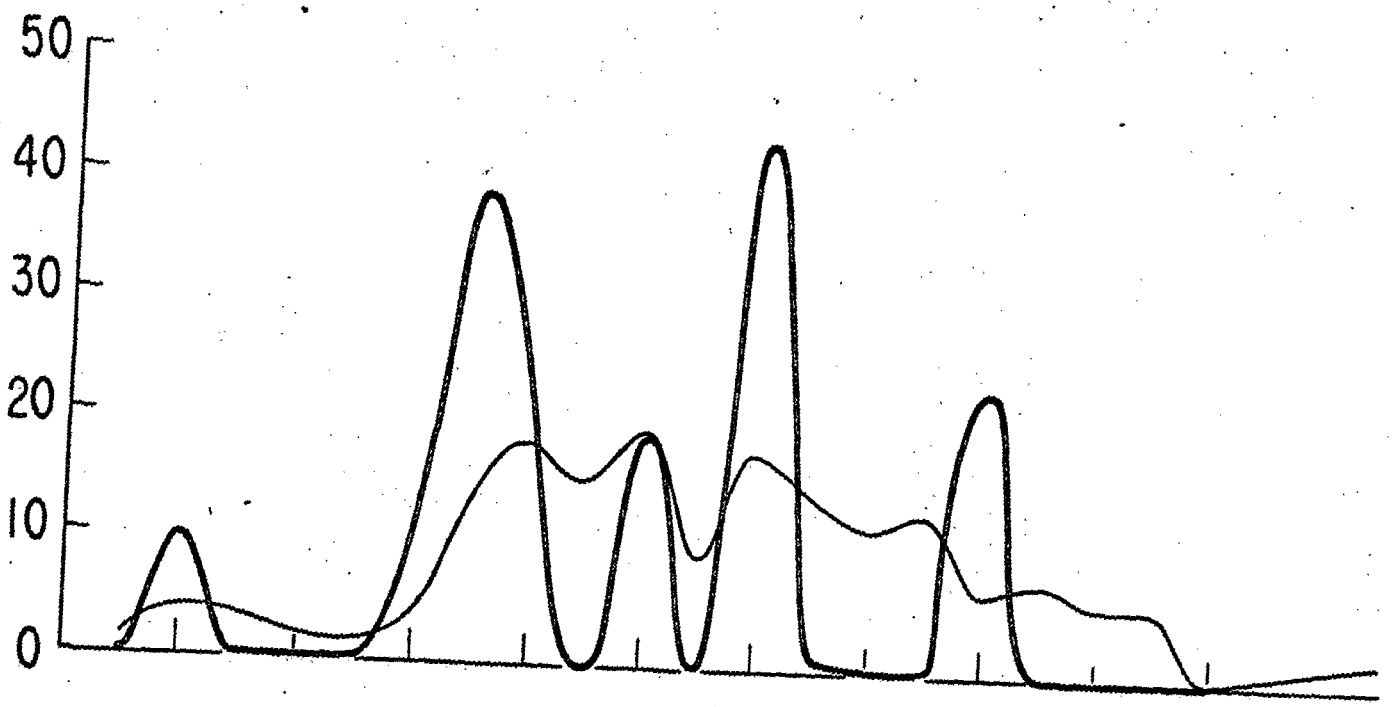

Fig. 4.4 True image (thick line) and data (thin line) after distortion by relative motion. $\sigma = 15$, $r = 0.2$, $I_0 = 3$, $\alpha = 10$, $D = 4$.

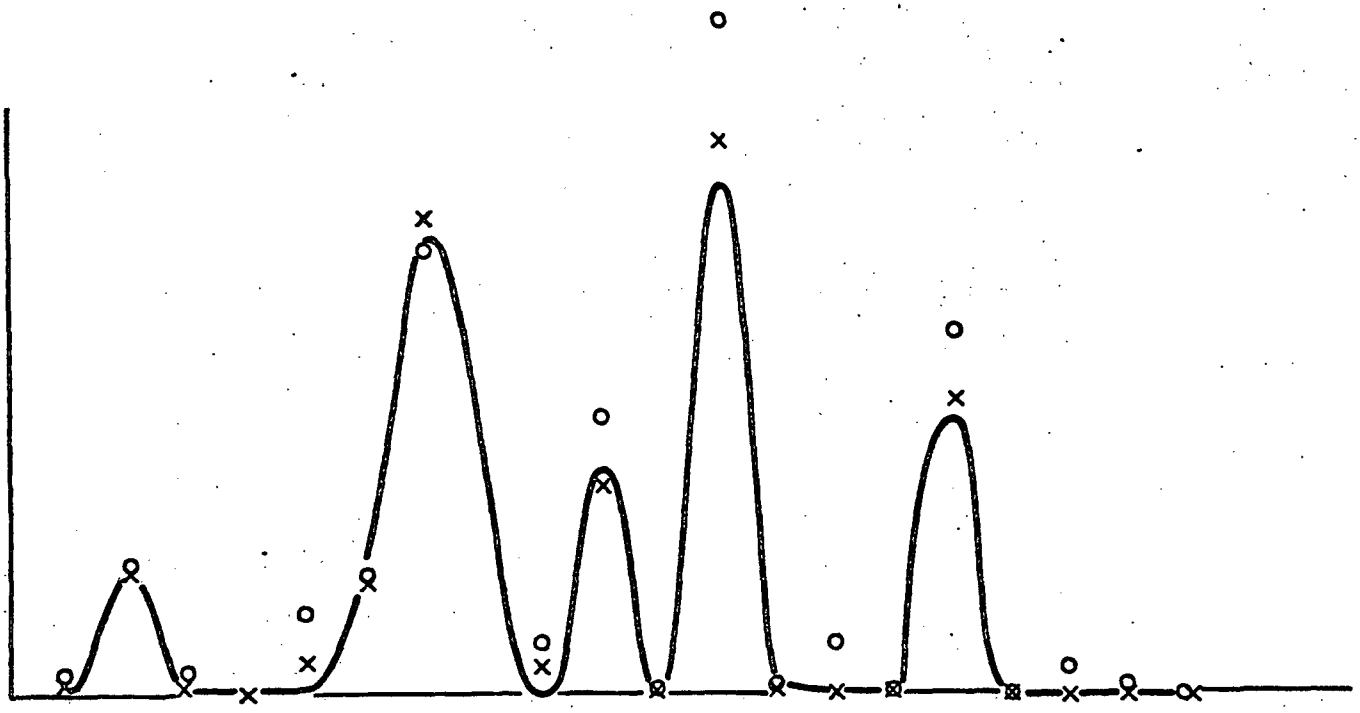

Fig. 4.5 Restoration of image in Fig. 4.4: × = maximum-likelihood; o = linear inverse filtering. $\sigma_b = \infty$.

Table II

$\sigma_b = 15, \quad r = 0.3, \quad I_0 = 5, \quad \alpha = 10$

Relative Squared Errors (percent)

	$\sigma_b = 15$, $r = 0.3$		$\sigma_b \to \infty$	
	linear	nonlinear	linear	nonlinear
1.	7.91	6.70	7.74	5.42
2.	9.69	7.07	8.11	5.70
3.	14.93	8.84	11.82	6.59
4.	13.57	7.23	15.79	6.90
5.	9.16	5.10	11.55	7.51
6.	13.94	8.12	18.44	9.73
* 7.	19.62	11.75	19.91	10.17
8.	9.51	6.86	9.94	10.74
9.	29.62	8.62	47.06	10.99
10.	16.04	14.55	16.81	14.27
11.	23.18	18.07	24.70	17.02
12.	22.19	12.78	25.73	17.05
mean	15.78	9.64	18.13	10.17
standard deviation	6.70	3.85	10.90	4.11

*Illustrated in Figs. 4.6 and 4.7

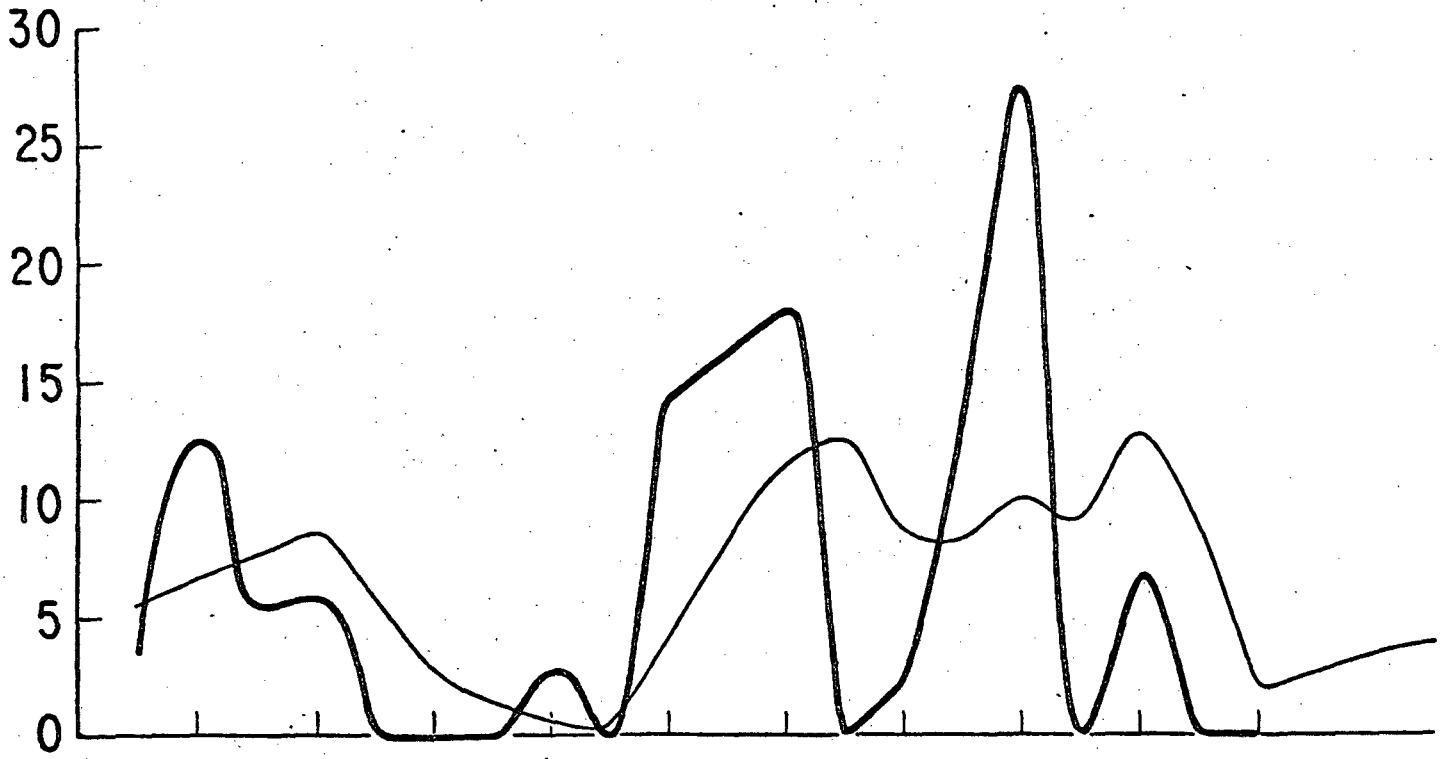

Fig. 4.6 True image (thick line) and data (thin line) after distortion by relative motion. $\sigma = 15$, $r = 0.3$, $I_0 = 5$, $\alpha = 10$, $D = 4$.

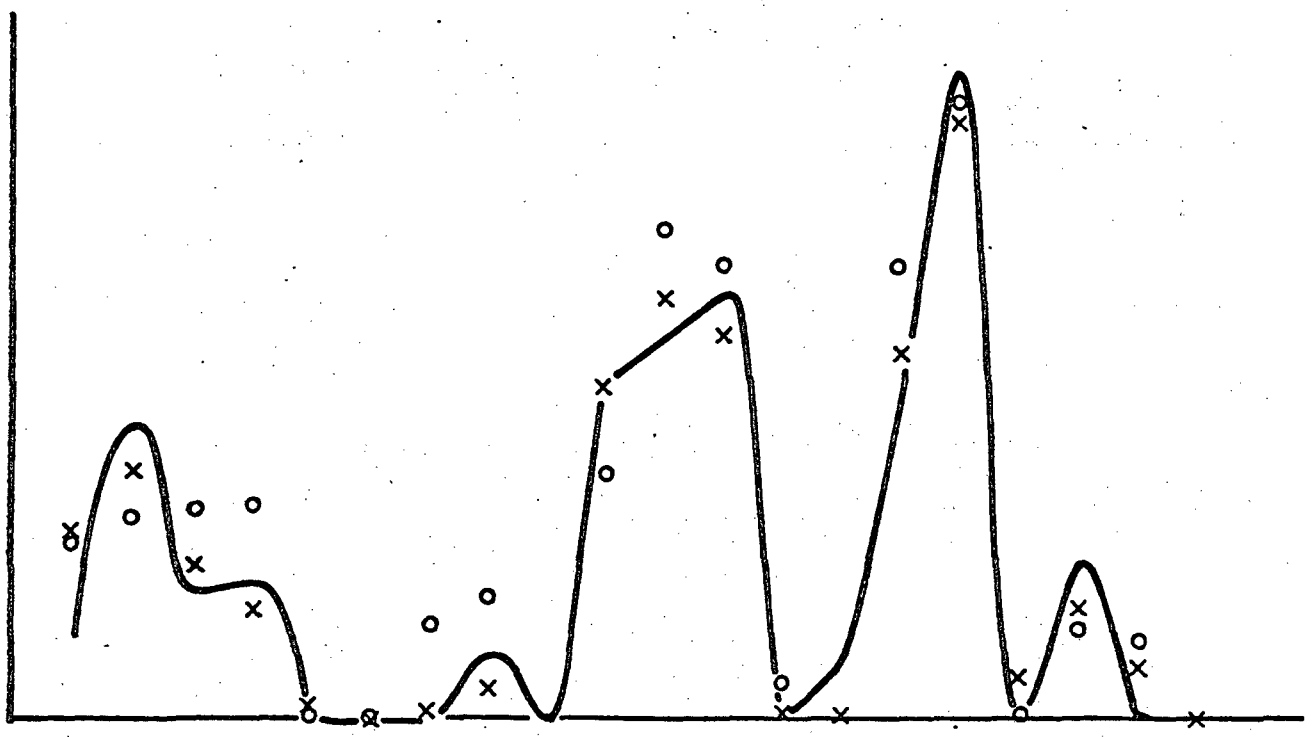

Fig. 4.7 Restoration of image in Fig. 4.6: × = maximum likelihood; o = linear inverse filtering, $\sigma_b = \infty$.

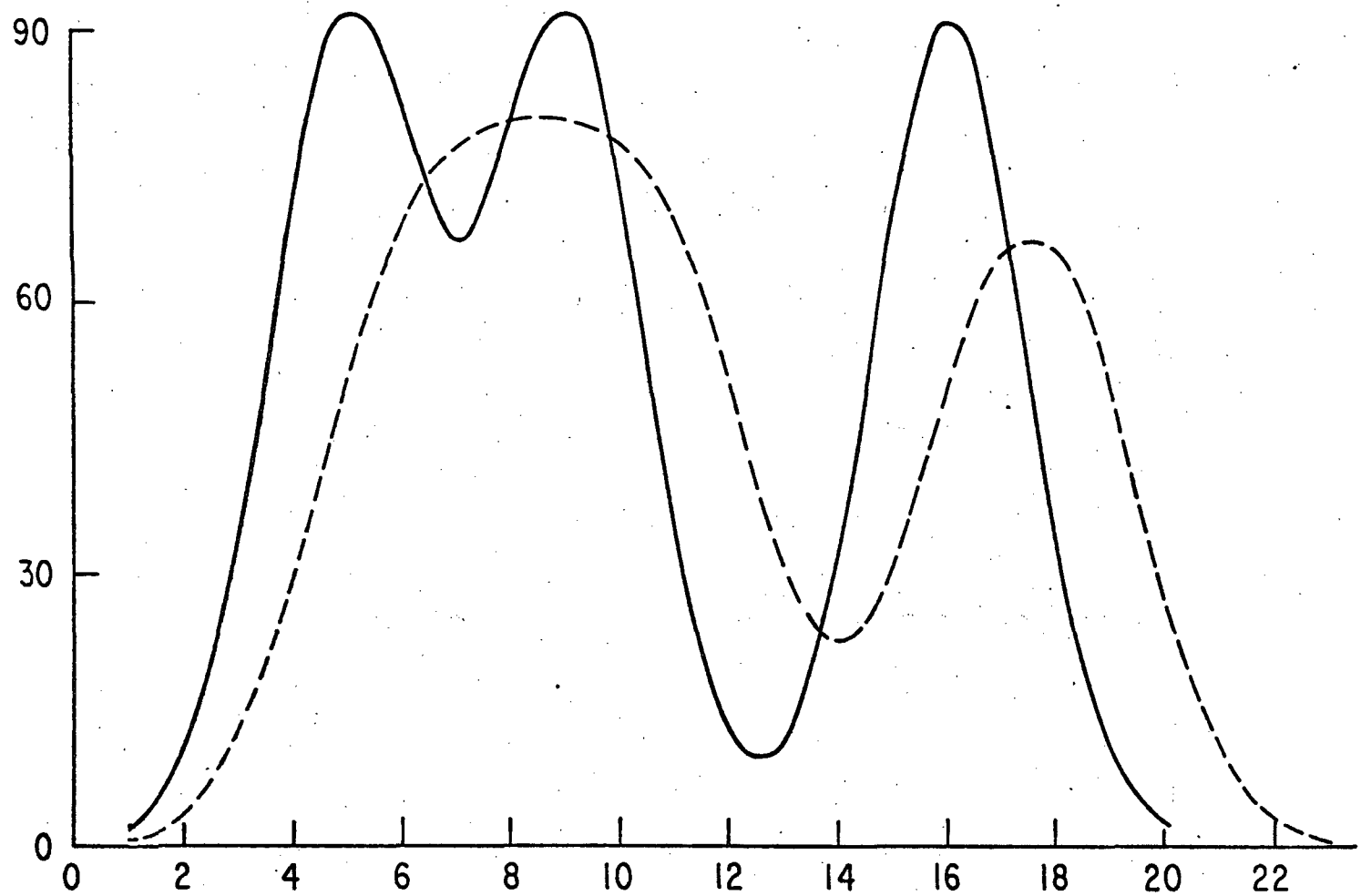

Fig. 4.8 Deterministic object of eq. (3.1) and its motion-blurred image; D = 4.

restored under the assumption that it belongs to the class of objects generated by the autoregressive process of (3.2) with $\sigma_b = 45$ and $r = 0.5$. The results of the maximum-likelihood and linear least-squaresestimates are shown by the crosses and circles in Fig. 4.9. In Figs. 4.10 and 4.11 we plot the relative squared errors obtained when different values of σ_b and r were used in the restoration procedure. For most of those values the maximum-likelihood estimate is the more accurate. It is also less sensitive to the choice of the values of σ_b and r than the linear least-squares estimate.

Diffraction was the second mode of image distortion considered. Figure 4.12 shows an autoregressive object generated with the values $\sigma_b = 15$ and $r = 0.2$ and distorted by the diffraction kernel of eq. (3.7) with W=1.2. The small central peak was eliminated by the diffraction, and we were unable to perceive it in any of the restorations. Data were generated from this image with $I_0 = 3$ and $\alpha = 10$. The restoration procedure assumed the values $\sigma = 20$ and $r = 0.2$, producing the points shown in Fig. 4.13. The three large peaks stand out in both restorations; the other two did not survive. In Fig. 4.14 are shown the percentage squared errors attained for various assumed values of the standard deviation σ_b and the correlation coefficient r. The method of inverse filtering failed completely for this object. Discarding prior knowledge ($\sigma_b = \infty$), the maximum-likelihood estimate yielded a relative squared error of 24.52%, which is roughly equal to the smallest value attained by any of the linear methods.

In Fig. 4.15, finally, we show the relative squared errors attained in restoring a diffracted Gaussian object of the form in eq. (3.1). The value of W in (3.7) was taken equal to 0.8, and $I_0 = 10$, $\alpha = 10$. In the restoration σ_b was taken equal to 15, and various values of r were tried. Over most of the range the maximum-likelihood estimator is much superior.

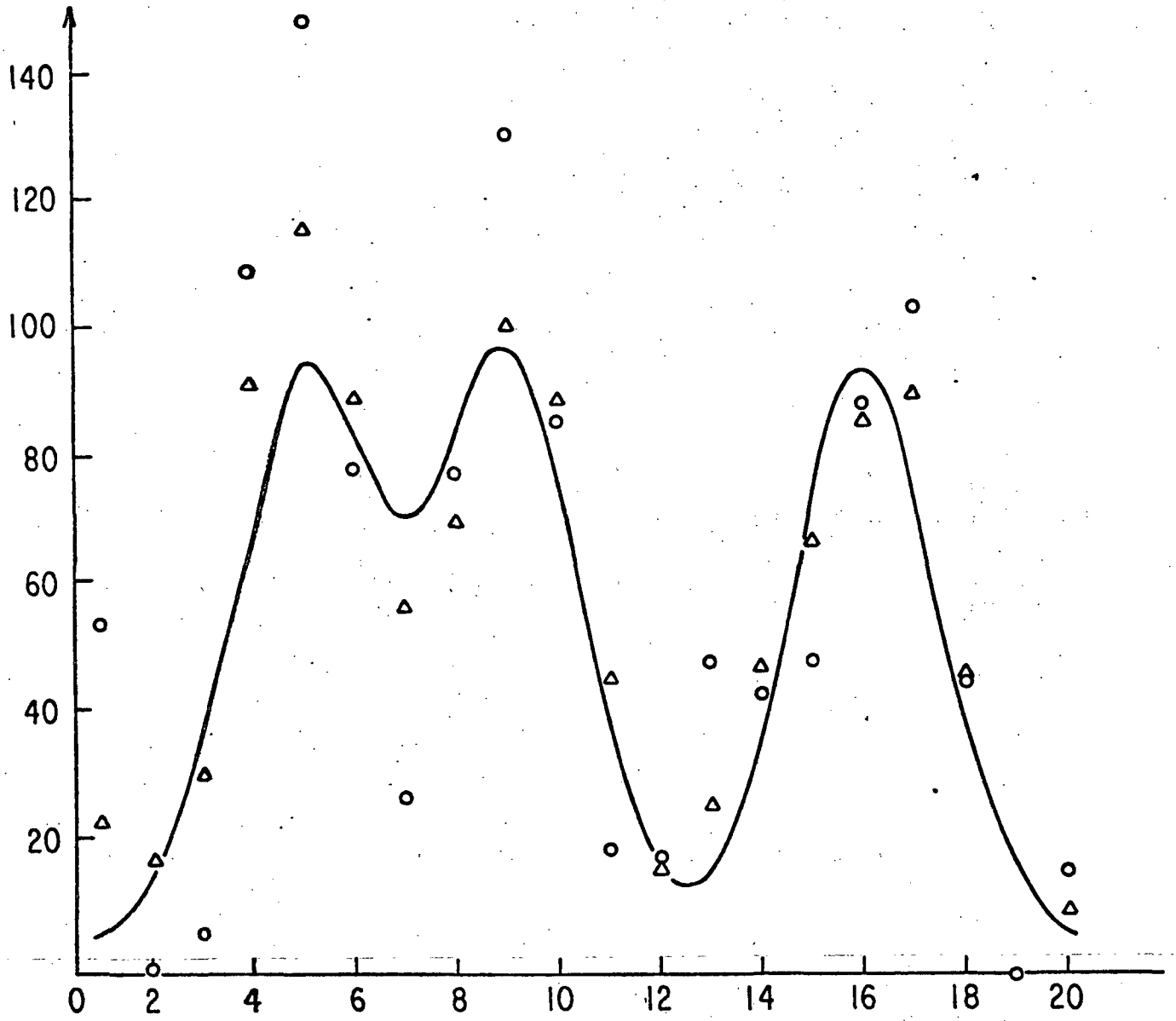

Fig. 4.9 Restoration of image in Fig. 4.8: $I_0 = 5$, $\alpha = 3$, $\sigma_b = 45$, $r = 0.5$. Δ = maximum-likelihood estimate, o = linear least-squares estimate.

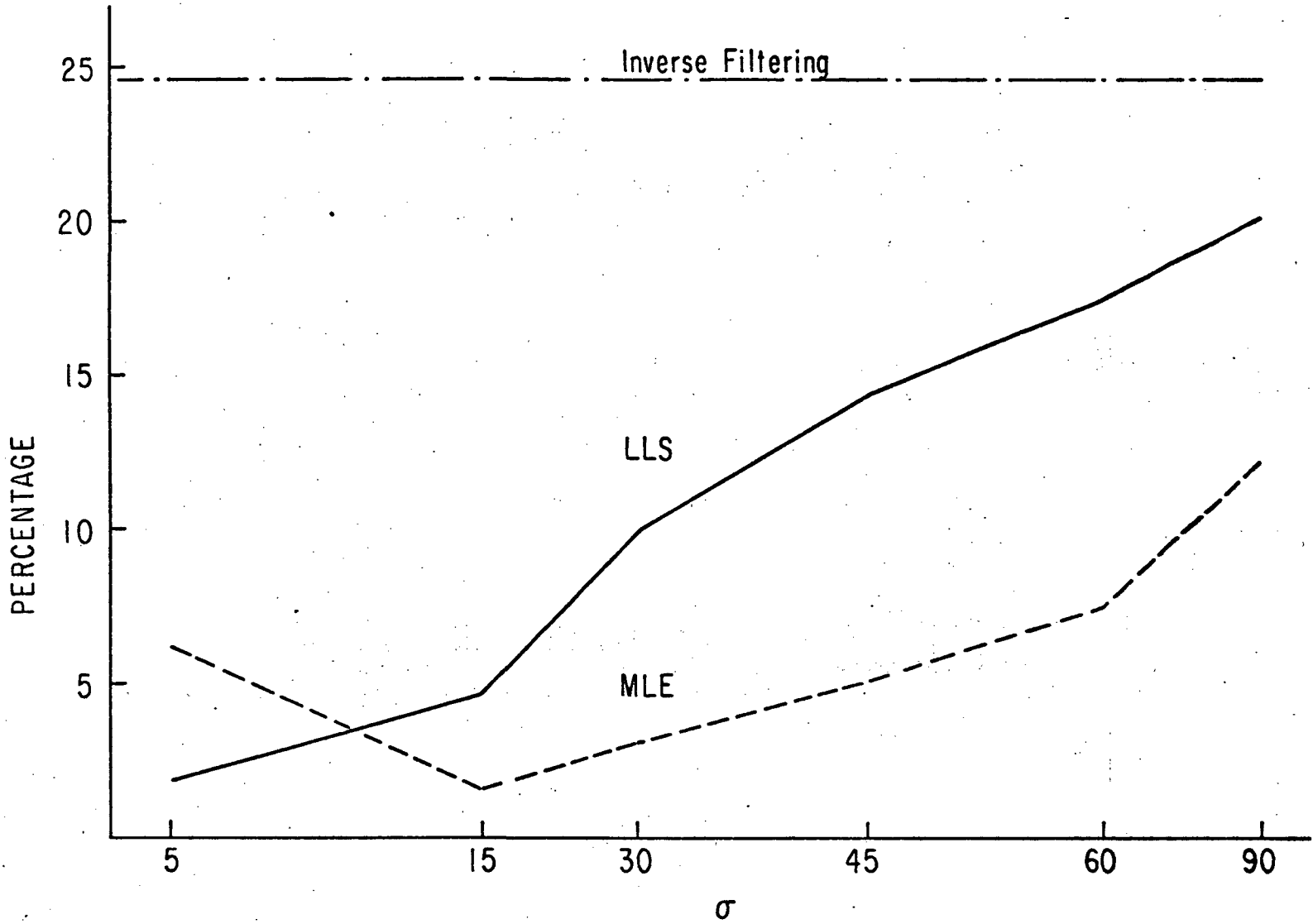

Fig. 4.10 Relative squared errors in restoring Fig. 4.8 for various assumed values of standard deviation σ_b; r = 0.5. Solid line-linear least squares; dashed line--maximum likelihood.

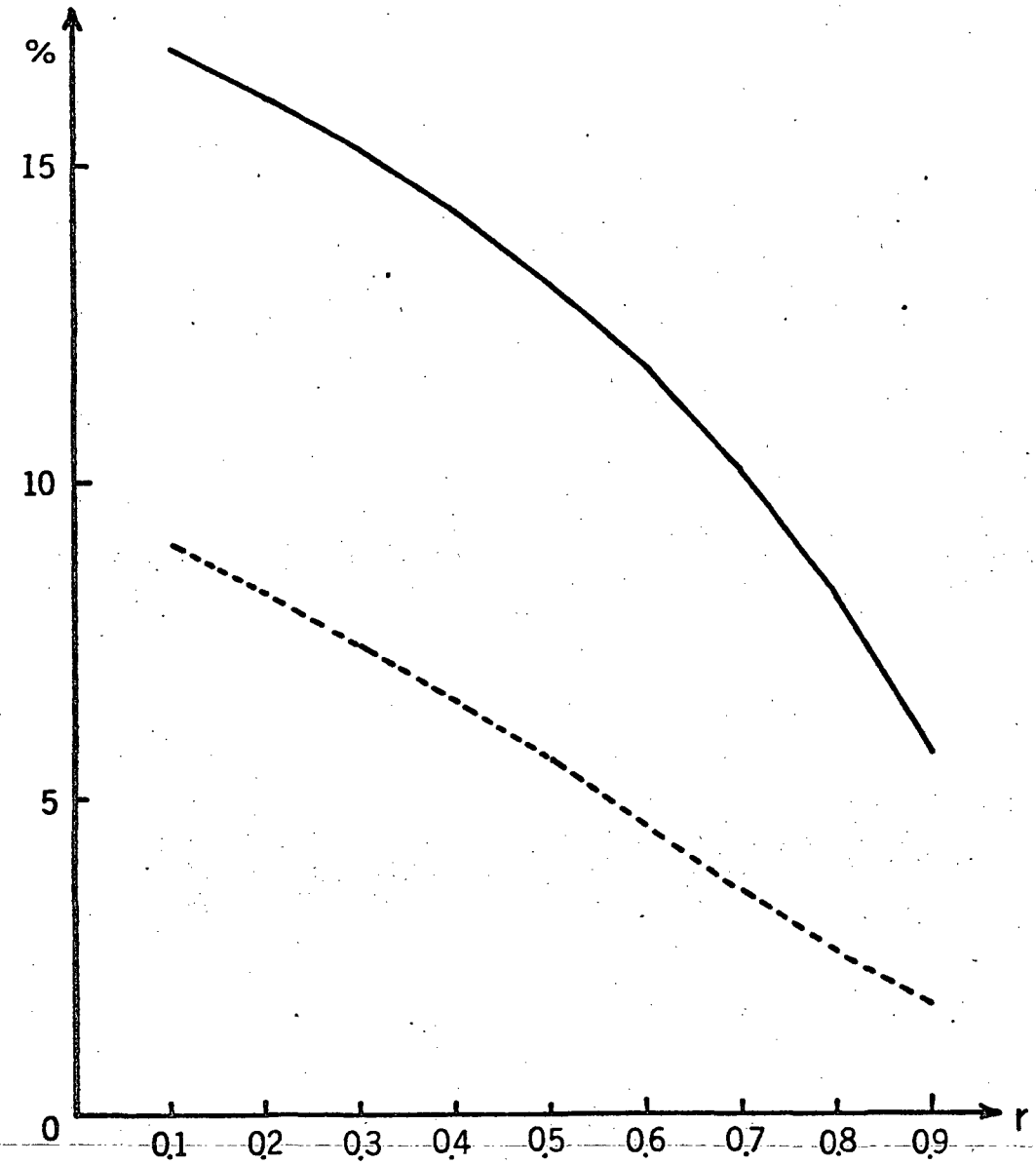

Fig. 4.11 Relative squared errors in restoring Fig. 4.8 for various assumed values of correlation coefficient r ; σ_b = 45. Solid line-linear least squares; dashed line-maximum likelihood.

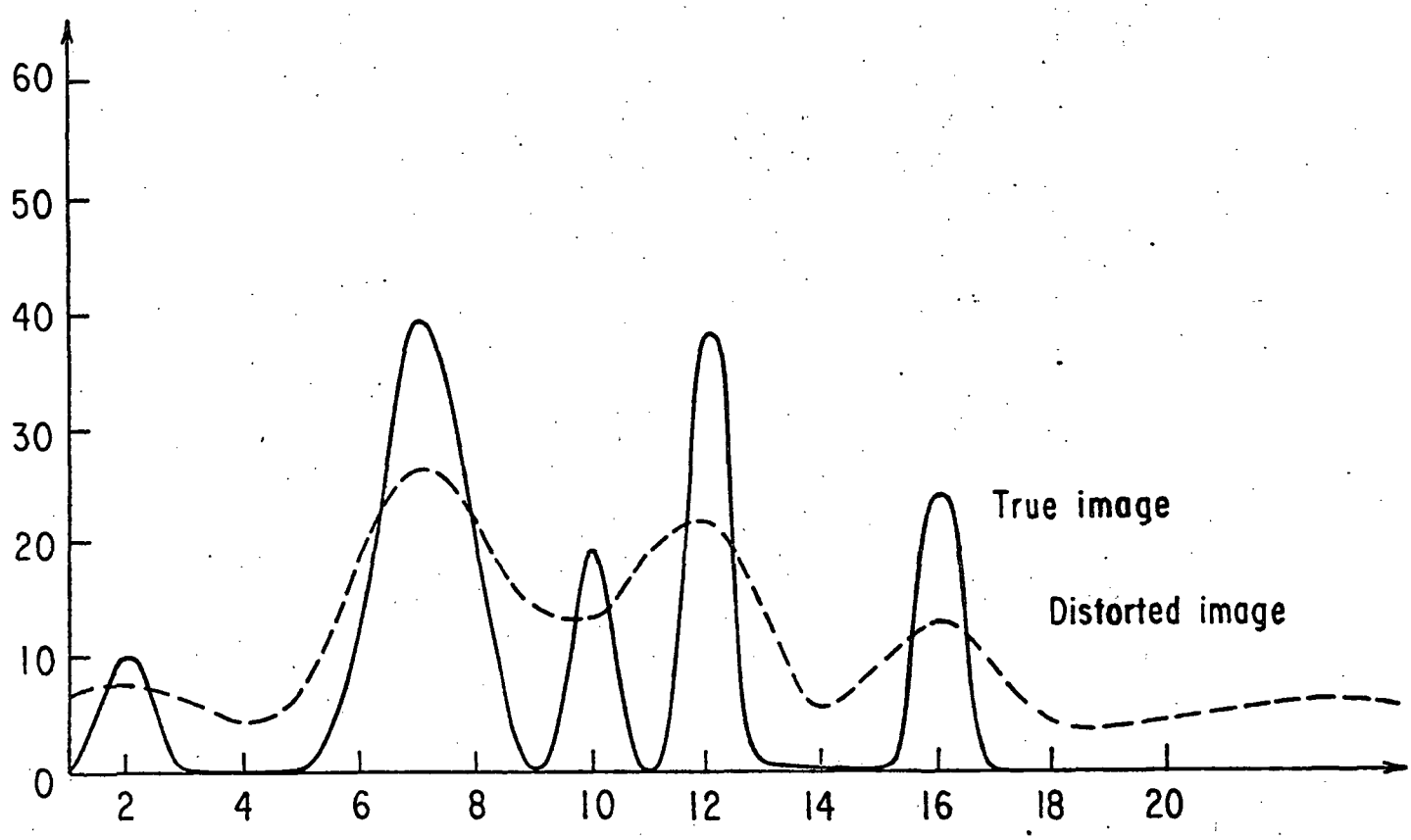

Fig. 4.12 Diffraction of an autoregressive image: $\sigma_b = 15$, $r = 0.2$, $I_0 = 3$, $\alpha = 10$, $W = 1.2$.

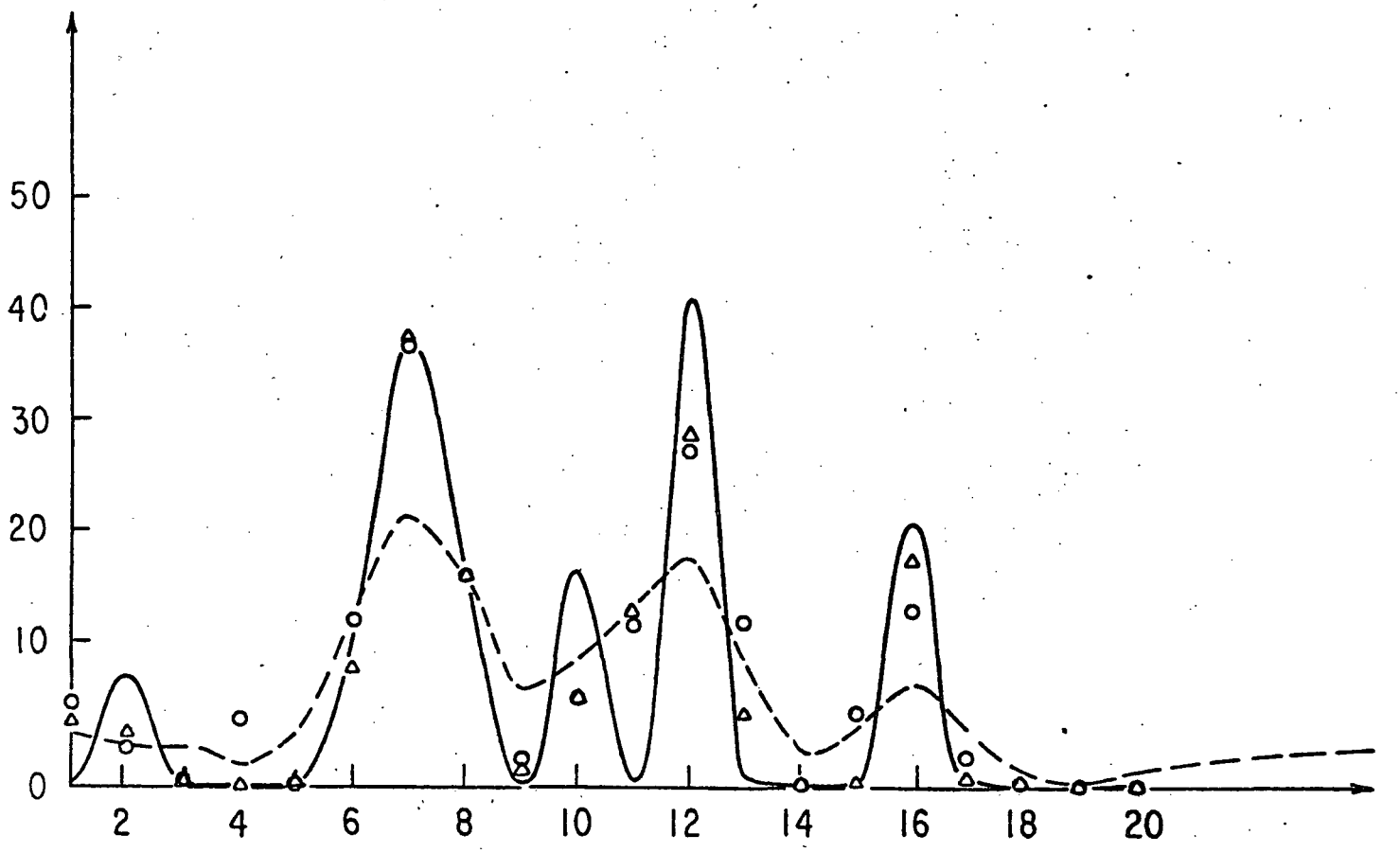

Fig. 4.13 Restoration of image in Fig. 4.12, taking $\sigma_b = 20$, $r = 0.2$, $W = 1.2$. o - linear least squares; Δ - maximum likelihood.

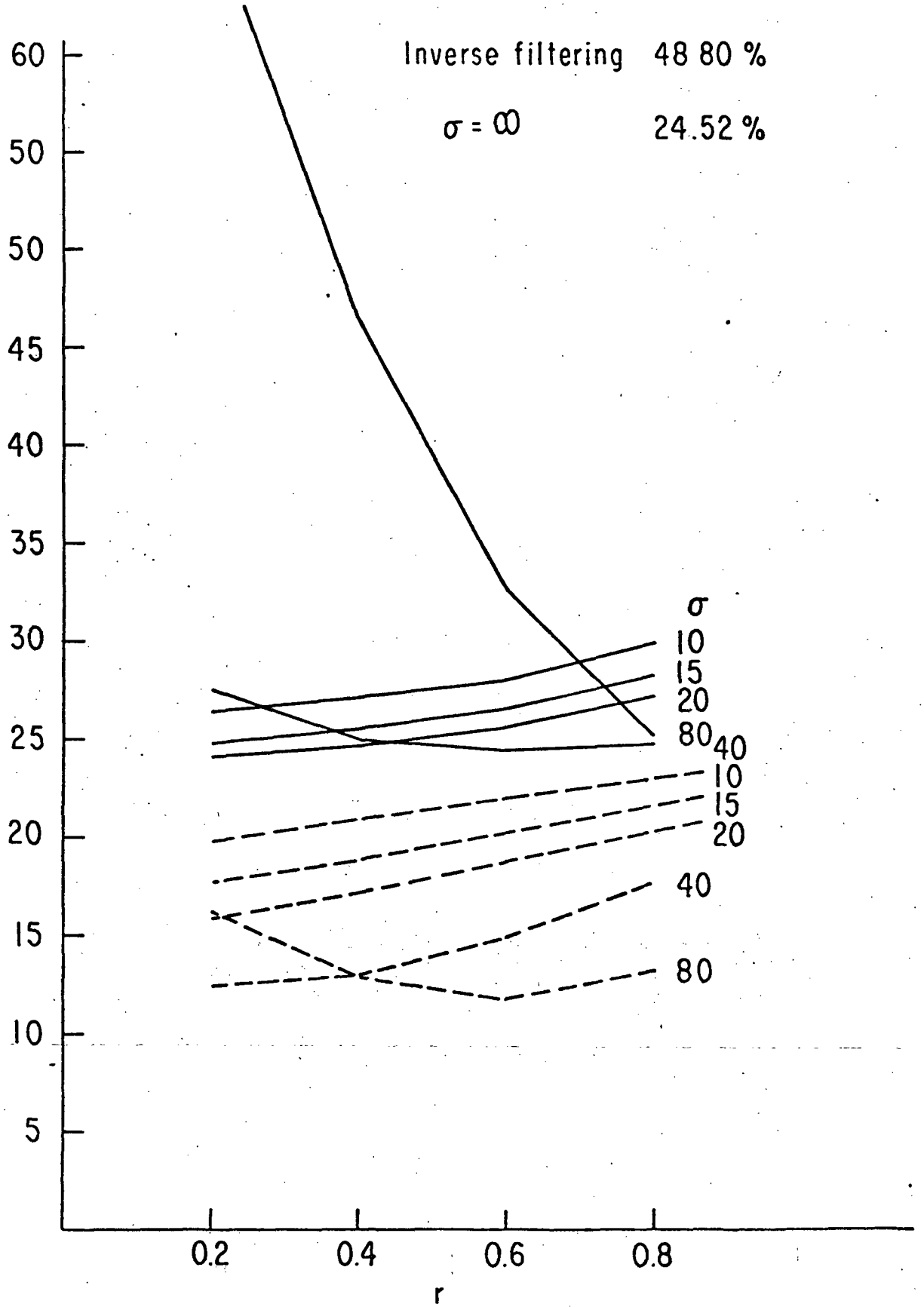

Fig. 4.14 Relative squared error (percent) as function of correlation coefficient r for restoring image in Fig. 4.12. Curves are indexed by the assumed standard deviation σ_b. Solid lines--linear least squares; dashed lines-maximum likelihood.

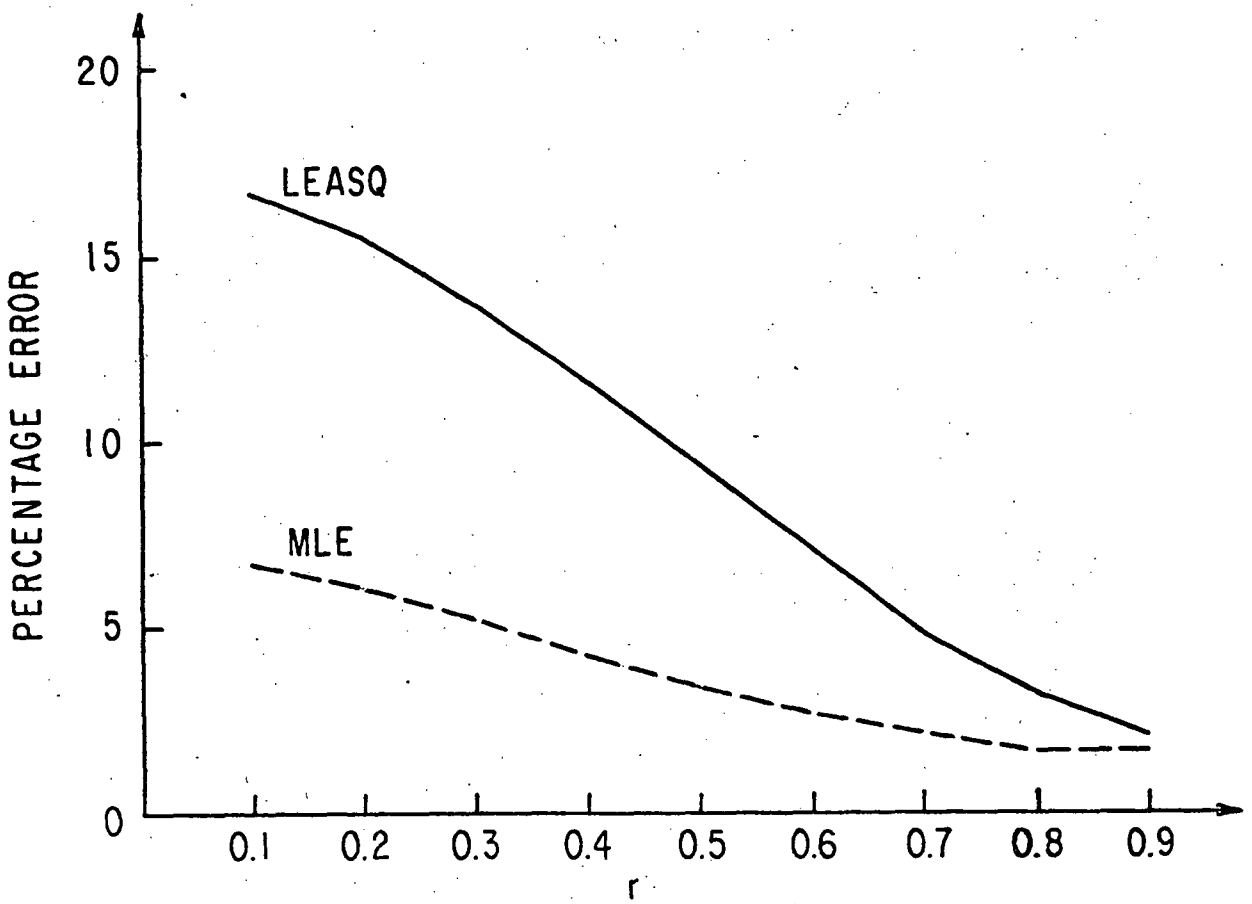

Fig. 4.15 Relative squared error (percent) in restoring image of eq. (3.1) after diffraction with W = 0.8 as function of assumed correlation coefficient r. $I_0 = 10$, $\alpha = 10$, $\sigma_b = 15$.

5. Conjugate Gradient Method

When the method of steepest descents nears the minimum value of the function $F\left(\left\{b_j\right\}\right)$, it has a tendency to oscillate between two points in the space of the estimanda, and this behavior protracts convergence. In an effort to find a faster computing routine, we turned to the method of conjugate gradients.[9] We applied it only to the function $F''\left(\left\{b_j\right\}\right)$ in (1.20), which is the negative logarithm of the likelihood function in the absence of prior knowledge about the class of objects. This method was unsuccessful.

Again a quadratic approximation to the function to be minimized is used, as in (2.1). The new trial point is derived from the original one at any stage by shifting in the direction $\underset{\sim}{d}$, which is one of the so-called"conjugate directions",

$$\underset{\sim}{b} = \underset{\sim}{b}^0 + \lambda \underset{\sim}{d}. \tag{5.1}$$

The value of λ is again chosen to minimize the new value of F",

$$\lambda = - \underset{\sim}{d} \cdot \underset{\sim}{g} \Big/ \underset{\sim}{d}^T \underset{\sim}{\underset{\sim}{A}} \underset{\sim}{d}, \tag{5.2}$$

where $\underset{\sim}{g}$ is the gradient vector whose components are given in (2.3), and $\underset{\sim}{\underset{\sim}{A}}$ is the matrix whose elements are a_{km} given in (2.7); in both cases we have set the inverse covariance matrix $\underset{\sim}{\underset{\sim}{\mu}}$ equal to $\underset{\sim}{\underset{\sim}{0}}$. The first direction $\underset{\sim}{d}$ is taken as $-\underset{\sim}{g}$; thereafter the direction $\underset{\sim}{d}$ for a stage is given in terms of the direction $\underset{\sim}{d}'$ for the previous stage by

$$\underset{\sim}{d} = \gamma \underset{\sim}{d}' - \underset{\sim}{g}, \tag{5.3}$$

where

$$\gamma = \underset{\sim}{g}^T \underset{\sim}{g} \Big/ \underset{\sim}{g}'^T \underset{\sim}{g}' \quad , \tag{5.4}$$

$\underset{\sim}{g}'$ being the gradient vector at the previous stage. After M such stages, $\underset{\sim}{d}$ is once again set equal to $-\underset{\sim}{g}$, the displacement of steepest descent, and the cycle is repeated.

The difficulties with this method apparently arose from the positivity constraint, which required our setting $b_j = -B_0 + \varepsilon$ whenever b_j fell below $-B_0$. Equation (5.1) then required us to change the corresponding component of the vector $\underset{\sim}{d}'$ to $\left(-B_0 - b_j^0\right)\Big/\lambda$ before calculating the new direction $\underset{\sim}{d}$ from (5.3). This occasionally caused the next move to be in an unfavorable direction, and the method did not converge. The images on which we tried it had high contrast, so that there were several adjacent points with zero illuminance, $B_0 + b_j = 0$. For an object of low contrast, with b_j never falling below $-B_0$, the conjugate-gradient method doubtless would have been superior, but for such objects the least-squares solution in (1.19) is generally adequate. The nonlinearity of the positivity constraint and the logarithmic singularity of the function $F''\left(\left\{b_j\right\}\right)$ to be minimized are too far from compatibility with the basic assumption of the conjugate-gradient method--that the function is a quadratic form—for it to be successful with objects of high contrast.

6. Conclusion

In the great majority of cases tried, the maximum-likelihood estimate of the true image was superior to that calculated by the linear least-squares method. The minimization of the negative logarithmic likelihood function by the method of steepest descents appears to be the most efficient procedure. The prior information about the class of images being restored had in most of our trials little influence on the outcome, and minimization of the function

$$F''\left(\{b_j\}\right) = \sum_{i=1}^{N} \left(\bar{n}_i - n_i \ell n\ \bar{n}_i\right) \tag{6.1}$$

with the constraints

$$\bar{n}_i = \alpha \left(I_0 + \sum_{j=1}^{M} K_{ij} b_j\right) , \tag{6.2}$$

$$b_j > -B_0 , \tag{6.3}$$

provided good estimates of the true images in most trials. This insensitivity to prior assumptions is gratifying, for the proper statistical description of a class of images is difficult to formulate.

This maximum-likelihood estimate is most appropriate when the image has a high contrast, with few or no photoelectrons being ejected from substantial areas of the picture. It is under such circumstances that the usual linear model of image formation with additive Gaussian noise is least accurate. The function in (6.1) takes the Poisson distributions of the data n_i into account, and the mechanism by which the image was distorted is embodied in the constraint (6.2). The third constraint (6.3) forces the restored image to have everywhere a non-negative illuminance. There remains to be investigated how the computation can

be made more efficient when applied to pictures furnishing large numbers of data.

References

1. J.L. Harris, Sr., "Image evaluation and restoration", J. Opt. Soc. Am. 56, 569-574 (May, 1966).

2. C.W. Helstrom, "Image restoration by the method of least squares", J. Opt. Soc. Am. 57, 297-303 (March, 1967).

3. D. Slepian, "Linear least-squares filtering of distorted images,", J. Opt. Soc. Am. 57, 918-922 (July, 1967).

4. B.R. Frieden, "Band-unlimited reconstruction of optical objects and spectra", J. Opt. Soc. Am. 57, 1013-1019 (August, 1967).

5. C.K. Rushforth, R.W. Harris, "Resolution, restoration, and noise", J. Opt. Soc. Am. 58, 539-545 (April, 1968).

6. C.W. Helstrom, "Linear restoration of incoherently radiating objects," J. Opt. Soc. Am. 62, 416-423 (March, 1972).

7. C.W. Helstrom, L. Wang, "Optimum detection of an optical image on a photoelectric surface," Trans. IEEE, AES-9 (July, 1973).

8. M.J.D. Powell, "A survey of numerical methods for unconstrained optimization," SIAM Review 12, 79-97 (January, 1970); reprinted in A.M. Geoffrion, Perspectives in Optimization, Reading, Mass.:Addison-Wesley (1972).

9. J. Kowalik, M.R. Osborne, Methods for Unconstrained Optimization Problems, New York: American-Elsevier (1968). See p. 41.

Figure Captions

Fig. 1.1 Image restoration system. The elements s of area ΔA emit numbers n_i of electrons that are processed to yield estimates $\hat{B} = B_0 + \hat{b}$ of samples of the true image of the radiance distribution B. O = object plane, A = aperture, I = image plane. A narrowband spectral filter for object and background light is not shown.

Fig. 3.1 True image (thick line) and the noisy degraded image (thin line) after blurring by relative motion. $I_0 = 3$, $\alpha = 10$, $\sigma = 25$, $r = 0.6$.

Fig. 4.1 True image of Fig. 3.1 and its restorations: o = linear estimate, Δ = inverse-filtering estimate, $\square$ = maximum-likelihood estimate.

Fig. 4.2 Percentage squared error in restoring Fig. 3.1 as function of assumed correlation coefficient r; curves are indexed by assumed standard deviation σ_b. Solid lines-linear estimate; dashed lines--maximum-likelihood estimate.

Fig. 4.3 Restoration of Fig. 3.1 without prior assumptions ($\sigma_b = \infty$). o = linear inverse filtering, × = maximum-likelihood.

Fig. 4.4 True image (thick line) and data (thin line) after distortion by relative motion. $\sigma = 15$, $r = 0.2$, $I_0 = 3$, $\alpha = 10$, $D = 4$.

Fig. 4.5 Restoration of image in Fig. 4.4: × = maximum-likelihood; o = linear inverse filtering. $\sigma_b = \infty$.

Fig. 4.6 True image (thick line) and data (thin line) after distortion by relative motion. $\sigma = 15$, $r = 0.3$, $I_0 = 5$, $\alpha = 10$, $D = 4$.

Fig. 4.7 Restoration of image in Fig. 4.6: × = maximum likelihood; o = linear inverse filtering, $\sigma_b = \infty$.

Fig. 4.8 Deterministic object of eq. (3.1) and its motion-blurred image; $D = 4$.

Fig. 4.9 Restoration of image in Fig. 4.8: $I_0 = 5$, $\alpha = 3$, $\sigma_b = 45$, $r = 0.5$. Δ = maximum-likelihood estimate, o = linear least-squares estimate.

Fig. 4.10 Relative squared errors in restoring Fig. 4.8 for various assumed values of standard deviation σ_b; $r = 0.5$. Solid line-linear least squares; dashed line--maximum likelihood.

Fig. 4.11 Relative squared errors in restoring Fig. 4.8 for various assumed values of correlation coefficient r ; $\sigma_b = 45$. Solid line-linear least squares; dashed line-maximum likelihood.

Fig. 4.12 Diffraction of an autoregressive image: $\sigma_b = 15$, $r = 0.2$, $I_0 = 3$, $\alpha = 10$, $W = 1.2$.

Fig. 4.13 Restoration of image in Fig. 4.12, taking $\sigma_b = 20$, $r = 0.2$, $W = 1.2$. o - linear least squares; Δ - maximum likelihood.

Fig. 4.14 Relative squared error (percent) as function of correlation coefficient r for restoring image in Fig. 4.12. Curves are indexed by the assumed standard deviation σ_b. Solid lines--linear least squares; dashed lines-maximum likelihood.

Fig. 4.15 Relative squared error (percent) in restoring image of eq. (3.1) after diffraction with $W = 0.8$ as function of assumed correlation coefficient r. $I_0 = 10$, $\alpha = 10$, $\sigma_b = 15$.